THE AMERICAN MATHEMATICAL MONTHLY

AN OFFICIAL JOURNAL OF

THE MATHEMATICAL ASSOCIATION OF AMERICA

(INCORPORATED)

DEVOTED TO THE INTERESTS OF COLLEGIATE MATHEMATICS

THE AMERICAN MATHEMATICAL MONTHLY, FOUNDED IN 1894 BY BENJAMIN F. FINKEL, WAS PUBLISHED BY HIM UNTIL 1913. FROM 1913 TO 1916 IT WAS OWNED AND PUBLISHED BY REPRESENTATIVES OF FOURTEEN UNIVERSITIES AND COLLEGES IN THE MIDDLE WEST

VOLUME 88

1981

PUBLISHED BY THE ASSOCIATION
MONTPELIER, VT, AND WASHINGTON, DC

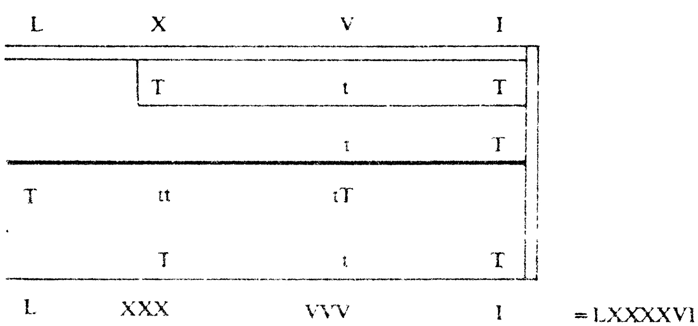
JANUARY

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 1

Banzhaf Index
Rigidity of Frameworks
Plücker Equations
Roman Numerals

Mathematics of Administration
Singular Monotone Functions
Computing π, e, $\sqrt{2}$
Teaching Symmetry



Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 1**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. (This is a change from the style previously used in this MONTHLY.) All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1979 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 18 months, Research Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201; NOTES, etc.: to the corresponding Associate Editor; REPRINT PERMISSION: to R. P. BOAS (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N.Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

FAIR APPORTIONMENT AND THE BANZHAF INDEX

BERNARD GROFMAN

School of Social Sciences, University of California, Irvine, CA 92717

1. Introduction. Justice Harlan's dissenting opinion in *Whitcomb* v. *Chavis* repudiated the Banzhaf index as a measure of voter power in apportionment decisions because of the supposed sensitivity of its calculations to minor variations in assumptions. Justice Harlan claimed such minor variations in assumptions could lead to differences in results on the order of magnitude of 120,000,000,000,000,000,000, but did not offer a rationale for his claim, the basis of which is not immediately obvious even to those familiar with the Banzhaf index and its properties. We provide formulae to determine Banzhaf power scores under varying assumptions as to voter partisan preferences and to verify Justice Harlan's calculations. We also briefly discuss the relevance of these formulae to a New York State Court of Appeals Decision, *Iannucci* v. *Board of Supervisors*, in which the Banzhaf index has been mandated as *the* test of the fairness of apportionment in county legislatures that make use of weighted voting.

2. Background. In three articles that appeared in American law journals in the mid-1960's, a lawyer named John Banzhaf III proposed to evaluate representation systems in terms of the extent to which they allocated "power" fairly [1], [2], [3]. Banzhaf's analysis makes use of game-theoretic notions in which power is equated with the ability to affect outcomes.

Consider a group of citizens choosing between two opposing candidates. To calculate the power of the individual voter, we generate the set of all possible voting coalitions among the district's electorate. If there are N voters in the district, then there will be 2^N possible coalitions. Then we ask, for each of these possible coalitions, whether a change in an individual voter's choice from candidate A to candidate B (or from candidate B to candidate A) would alter the electoral outcome. If so, that voter's ballot is said to be *decisive*. The (absolute) Banzhaf index of a voter's power is defined as the number of the voter's decisive votes divided by 2^N. The higher the percentage of voter coalitions in which a voter's vote is decisive, the higher that voter's power score. The Banzhaf index has considerable intuitive appeal; power is based on ability to affect outcome.

For single-member district systems (smds) whose districts are of equal population, all voters have identical power. But what about the case of multiple-member district systems (mmds), with districts of more than one size? Here, since the voters who elect k representatives have k times as much impact as voters who can elect only one representative, we might think that to equalize voter power we should assign to each district a number of representatives proportional to the size of the district's population since, intuitively, we would expect a voter's ability to decisively affect outcomes should be inversely proportional to district population. Banzhaf [2] pointed out that this argument is mathematically incorrect.

In a two-candidate/party contest where all voters have equal weight, in order for a voter to be decisive in a district of size N the rest of the voters (who are $N-1$ in number) must split half for one candidate/party and half against. A straightforward combinatoric analysis reveals ([2], [7], [6]; *Whitcomb* v. *Chavis* (1970) 403 U.S. at 145 n. 23) that, *if all combinations of vote outcomes are equally likely* (i.e., if each voter is equally likely to vote for either candidate/party), then the number of each member's decisive votes, b, is given by:

$$b=\frac{2(N-1)!}{\left(\frac{N-1}{2}\right)!\left(\frac{N-1}{2}\right)!}. \tag{1}$$

Bernard Grofman is Professor of Political Science and Social Psychology, School of Social Sciences, University of California, Irvine. He is a specialist in mathematical models of collective decision-making, with numerous published articles on such topics as jury verdict choice, voter turnout, coalition formation models, and the political consequences of electoral laws.—*Editors*

We can examine the link between b and N by using Stirling's approximation [4], [2], [7], [6],

$$N! \approx e^{-N} N^{N} (2\pi N)^{1/2} \tag{2}$$

to rewrite (1) as

$$b \approx \frac{2^{N+1}}{(2\pi(N-1))^{1/2}}. \tag{3}$$

Thus, each member's Banzhaf index, which we shall denote by B_i, is simply

$$B_i = \frac{2\binom{N-1}{\frac{1}{2}(N-1)}}{2^N} \approx \frac{2}{(2\pi(N-1))^{1/2}}. \tag{4}$$

This analysis can be applied to electoral systems involving both single- and multi-member districts. We see from expression (4) that B_i is approximately inversely proportional to the square root of N district population. Thus, if we wish to assign all voters equal power to affect outcomes, we should assign each district a number of representatives proportional to the *square root* of district population, rather than directly proportional to district population.[1]

In a case decided in 1970, *Whitcomb* v. *Chavis*, 403 U.S. 143, the Supreme Court dealt directly with Banzhaf's concept of voter power. The case involved Indiana's scheme of single and multiple member districts for its state legislature. The plaintiffs, citing Banzhaf's work [2], argued that voters in the multiple member districts were overrepresented because citizens in the larger district had a power disproportionate to their population.

We can make this argument explicit as follows: if mmds (with population mN) elected m representatives, then each voter in such a district would have a power proportionate to $m/\sqrt{mN}$; while those in smds with a population of N would have a power proportional to $1/\sqrt{N}$. Since $m/\sqrt{m} > 1$, for $m > 1$, this would be denying to all citizens an "equally effective voice in the election of members of his legislature" (377 U.S. at 565).

The Banzhaf [2] argument was rejected decisively in *Whitcomb*, both in the majority opinion and in Justice Harlan's dissenting opinion. Only Harlan's opinion, however, dealt forthrightly with the intellectual merits of the Banzhaf argument.

In Justice Harlan's dissenting opinion in *Whitcomb* v. *Chavis* he pokes some more or less good-natured fun both at the Banzhaf index and at his brethren on the Court. First, Harlan (403 U.S. at 168, n. 2) cites the majority views on the Banzhaf index; to wit that, while mathematically correct, its implications can safely be ignored because it "does not take into account any political or other factors which might affect the actual voting power of the residents, which might include party affiliation, race, previous voting characteristics or any other factors which go into the entire political voting situation" (Ante at 145, 146). Then he retorts sarcastically that "precisely the same criticism applies with even greater force to the one man, one vote opinions of the Court. *The only relevant difference between the elementary arithmetic on which the Court relies and the elementary probability theory on which Professor Banzhaf relies is that calculation in the latter field can't be done on one's fingers*" (403 U.S. at 168 n. 2 [emphasis ours]).

Harlan then goes on to lampoon the absurdity of the Banzhaf index's simplifying assumptions e.g., that "the voting habits of all members of the electorate are alike" and that "each voter is equally likely to vote for either candidate before him" (403 U.S. at 168), and asserts that "minor variations in these assumptions can lead to major variations in results" (403 U.S. at 169). Harlan looks at a case involving 300,000 voters and claims:

> If the temper of the electorate changes by one-half of one percent (more precisely, the result follows if the second of Professor Banzhaf's assumptions is altered so that the probability of each voter's selecting candidate A over candidate B is 50.5% rather than 50%) then each individual's voting power is reduced by a factor of approximately 1,000,000. Or if a few of the 300,000 voters are committed—say 15,000 to candidate A and 10,000 to candidate B—the probability of any individual's casting a tie-breaking vote is reduced by a factor on the rough

order of 120,000,000,000,000,000,000. Obviously in comparison with the astronomical differences in voting power which can result from such minor variations in political characteristics, the effects of the 12% and 28% population variations considered in *Abate* vs. *Mundt* and in this case are *de minimis*.

Harlan does not indicate how he arrives at the figures 1,000,000 and 120,000,000,000,000,000,000. It is far from obvious, even to someone familiar with the Banzhaf index and its properties, where those numbers come from.

In this paper we shall investigate the sensitivity of Banzhaf power calculations to variations in assumptions about coalitional probabilities to see how results can vary as radically as Harlan claims they do. This issue is an important one, not merely because it will clear up the puzzlement of scholars reading Harlan's opinion in *Whitcomb* who are unable to determine on what basis his calculations are derived, but also because the Banzhaf index is enshrined into law as *the* test of fair apportionment of *weighted* voting systems in the state of New York. Thus, the properties of the Banzhaf index remain very much of interest, despite the rejection in *Whitcomb* of its applicability to the mixed single- and multi-member district case.[2, 3]

3. The Impact of Varying the Assumptions on Which Banzhaf Power Calculations Are Based. First, let us consider the case where there are N identical voters, each of whom votes for candidate A with probability p (where p need not equal $\frac{1}{2}$). A given voter is decisive when the remaining $N-1$ voters divide equally between the two candidates. For specified p the expected proportion of coalitions in which the voter will be decisive, which we shall call $B_{i_{(p)}}$, is given by

$$B_{i_{(p)}}=\frac{2\binom{N-1}{\frac{1}{2}(N-1)}p^{\frac{1}{2}(N-1)}(1-p)^{\frac{1}{2}(N-1)}}{2^N}. \tag{5}$$

The impact of a deviation from equiprobability in p varies as a function of N. In particular, the ratio of $B_{i_{(p)}}$ to $B_{i_{(1/2)}}$ is given by

$$\frac{B_{i_{(p)}}}{B_{i_{(1/2)}}}=(2p)^{\frac{1}{2}(N-1)}[2(1-p)]^{\frac{1}{2}(N-1)}. \tag{6}$$

To simplify calculation of this ratio it is convenient to take logarithms. Thus

$$\frac{B_{i_{(p)}}}{B_{i_{(1/2)}}}=\operatorname{antilog}\left[\tfrac{1}{2}(N-1)(\log 2p+\log(2(1-p)))\right]. \tag{7}$$

For the 300,001 voter example considered by Harlan, for which $p=.505$, the ratio given in expression (7) is 3.1×10^{-7}.

Now, let us consider the case where N voters, of whom $K_1\left(\frac{1}{2}(N+1)>K_1\right)$ are committed to candidate A and $K_2\left(\frac{1}{2}(N+1)>K_2\right)$ to candidate B. For simplicity, let us initially assume $p=\frac{1}{2}$, i.e., that all coalitions of uncommitted voters are equally likely. A voter is decisive only when $\left(\frac{1}{2}(N-1)-K_1\right)$ of the (N_1-K_2-K-1) remaining undecided voters vote for A, i.e., when $\left(\frac{1}{2}(N-1)-K_2\right)$ of the remaining $(N-K_1-K_2-1)$ undecided voters vote for candidate B. There are $2^{N-K_1-K_2}$ possible voter coalitions, since K_1+K_2 of the voters are committed voters. The proportion of feasible coalitions in which the given voter will be decisive is given by

$$\frac{2\binom{N-K_1-K_2-1}{\frac{1}{2}(N-1)-K_1}}{2^{N-K_1-K_2}}. \tag{8}$$

Using Stirling's formula, and if we let $x=N-K_1-K_2-1$, $y=(N-1-2K_1)/2$, and $z=(N-1-2K_2)/2$, expression (8) can be approximated by

$$\frac{2e^{-x}x^x(2\pi(x))^{1/2}}{e^{-y-z}y^yz^{1/2(N+1-K_1-K_2)}\cdot 2\pi\sqrt{yz}\cdot 2^{N-K_1-K_2}}. \tag{9}$$

Using the approximations given in expression (9), the ratio of expression (8) to expression (4) can, after some algebra, be expressed as

$$\frac{((N-1)x)^{1/2}x^x}{2(yz)^{1/2}(2y)^y(2z)^z}. \tag{10}$$

In the case that $K_1, K_2 \ll N$ (read K_1, K_2 considerably less than N), we may approximate expression (10) by

$$\frac{\left(\frac{N-1}{x}\right)^{1/2}x^x}{(2y)^y(2z)^z}. \tag{11}$$

By expressing the formulae of (10) or (11) in terms of logarithms (cf. expression (7)) calculation of numerical results is straightforward. For the case where $N=300{,}001$, $K_1=15{,}000$, and $K_2=10{,}000$, we obtain a value of roughly 10^{-20}, as per Harlan's assertion.

In the special case that $K_1=K_2=K$, expression (9) directly simplifies to

$$\left(\frac{N-1}{N-1-2K}\right)^{1/2}. \tag{12}$$

Note also that for K_1, $K_2 \ll N$ we have

$$\left(\frac{N-1}{N-1-K_1-K_2}\right)^{1/2} \approx 1. \tag{13}$$

4. Conclusions. We have provided formulae that specify the impact on Banzhaf power calculations of deviations from the assumption of equiprobable coalitions and identical voters; and we have vindicated Justice Harlan's mathematics.

Acknowledgments. This research was supported by NSF Grant Soc. 77−24474, Political Science Program. It represents a continuation of work done jointly with Professor Howard Scarrow, State University of New York at Stony Brook, and came about because Professor Scarrow asked me how Justice Harlan came up with the figure of 120,000,000,000,000,000,000 given in *Whitcomb* v. *Chavis* (1976) 403 U.S. at 169. This paper was written to answer that question—a question that I'm sure has been asked (but rarely, if at all, answered) by everyone who has ever read Harlan's dissenting opinion in the *Whitcomb* case. I am indebted to the staff of the Word Processing Center of the School of Social Sciences, University of California, Irvine, for translating my scribbles and hand-written mathematical formulae into typed copy.

Notes

1. Such an assignment may violate other norms. If, for example, we assigned one representative for every 100 population in the square root of district size, then if there are 20,000 population spread equally over 2 smds, these voters (10,000 per district) would be entitled to have 2 representatives, 1 per district, since the square root of 10,000 is 100. Similarly, if there are 40,000 citizens spread equally over 4 smds (10,000 each) they would be entitled to 4 representatives. However, a single mmd of size 40,000 would be allocated only 2 representatives, since the square root of 40,000 is only 200. Thus, in this example, 20,000 voters would be entitled to as many representatives as 40,000 voters, the allocation of representatives to the 40,000 voters depends on how voters are divided among the districts.

2. The New York Court of Appeals in *Iannucci* v. *Board of Supervisors of the County of Washington* (1967) 282 N.Y.S. 2d 502 held:

> The principle of one man–one vote is violated when the power of a representative to affect the passage of legislation by his vote. . . does not roughly correspond to the proportion of the population in his constituency. Thus, for example, a particular weighted voting scheme would be invalid if 60% of the population were represented by a single legislator who was entitled to cast 60% of the votes. Although his vote would apparently be weighted only in proportion to the population he represented, he would actually possess 100% of the voting power whenever a simple majority was all that was necessary to enact legislation.

Similarly a plan would be invalid if it was *mathematically impossible* for a particular legislator representing say 5% of the population to ever cast a decisive vote. Ideally, in any weighted voting plan, it should be mathematically possible for every member of the legislative body to cast the decisive vote on legislation in the same ratio which the population of his constituency bears to the total population. Only then would a member representing 5% of the population have, at least in theory, the same voting power (5%) under a weighted voting plan as he would have in a legislative body which did not use weighted voting—e.g., as a member of a 20-member body with each member entitled to cast a single vote. This is what is meant by the one man–one vote principle as applied to weighted voting plans for municipal governments. A legislator's voting power, measured by the mathematical possibility of his casting a decisive vote, must approximate the power he would have in a legislature which did not employ weighted voting.

The *Iannucci* decision has had tremendous impact on New York County government, where 24 of 57 counties now use some form of weighted voting. In *Iannucci* the court held that counties would have to submit computer calculations of Banzhaf index scores to verify that any proposed legislature weight satisfied the *Iannucci* guidelines. Since weights assigned directly proportional to population represented often resulted in Banzhaf scores discrepant with weights (e.g., in a 3-member legislature with districts of size 2,000, 2,000, and 1,000 and a weight assignment of 2, 2, and 1, the Banzhaf scores of all 3 legislators are identical, despite the fact that one represents only half as many voters as the other two and has only half the weight they do), this has necessitated some counties' hiring professional assistance to generate weights that yield power scores concordant with population. Lee Papayanopoulous, a mathematician and computer programmer, has provided this service for well over a dozen New York counties in the past decade. (For a further discussion of the *Iannucci* decision and its effects on representation systems in the state of New York, see [5].)

3. The formulae we provide are relevant to challenges to the appropriateness of the Banzhaf index as a measure of *legislator* power, since in the legislature *even more than in the electorate* the assumptions of identical actors and equiprobable coalitions seem far indeed from the realities of politics; however, they must be adapted to deal with relative (normalized) Banzhaf power scores rather than the absolute (non-normalized) power scores discussed above. Moreover, for legislators, a smaller N leads to *far less* extreme variations in power as we vary coalitional probabilities.

Furthermore, unlike the U.S. Supreme Court in *Whitcomb*, which rejected Banzhaf calculations because of their lack of political realism, the New York Court of Appeals in *Iannucci* thought that this divorce from political realities was a positive feature of the Banzhaf index! In *Iannucci* actual voting patterns were held to be irrelevant. The sole criterion to be used in determining the constitutionality of a weighted voting scheme "is the mathematical voting power that each legislator possesses in theory—i.e., the indicia of representation—and not the actual voting power he possesses in fact—i.e., the indicia of influence" (20 N.Y. 2d at 252).

References

1. John F. Banzhaf III, Weighted voting doesn't work: A mathematical analysis, Rutgers Law Review, 13 (1965) 317–343.

2. ______, Multi-member electoral districts—do they violate the "one-man-one-vote" principle, Yale Law Journal, 75 (1966) 1309–1388.

3. ______, One man, 3,312 votes: A mathematical analysis of the Electoral College, Villanova Law Review, 13 (Winter 1968) 314–322 (see also comments on pp. 333–346).

4. William Feller, Introduction to Probability Theory, vol. 1, rev. ed., John Wiley, New York, 1971.

5. Bernard Grofman and Howard Scarrow, *Iannucci* and its aftermath: Game theory and weighted voting in the state of New York, in Steven Brams (ed.), Game Theory, Physica-Verlag, Vienna, 1979.

6. William F. Lucas, Measuring Power in Weighted Voting Systems, Technical Report No. 227, Department of Operations Research, Cornell University, Ithaca, 1974.

7. E. Walther, An Analysis of Weighted Voting Systems Using the Banzhaf Value, Technical Report No. 809, School of Operations Research and Industrial Engineering, Cornell University, Ithaca.

MISCELLANEA

49. Who said *that*?

1. It will not occur to anyone to undervalue the merit of the mathematicians.
2. Mathematics talks about the things which are of no concern at all to man.

(Answers on p. 52.)

RIGID AND FLEXIBLE FRAMEWORKS

B. ROTH

Department of Mathematics, University of Wyoming, Laramie, WY 82071

1. Introduction. Consider a triangle or a square in the plane $\mathbb{R}^2$ whose edges are inextendible, incompressible rods which are joined but rotate freely at the vertices. The square is said to be flexible in $\mathbb{R}^2$ since the square can move continuously (or fall over) into a family of rhombi, as shown in Fig. 1. However, the triangle is said to be rigid in $\mathbb{R}^2$ since the three rods determine the relative positions of the three vertices. Similarly, a tetrahedron in $\mathbb{R}^3$ consisting of six rods, connected but freely pivoting at the four vertices, is rigid in $\mathbb{R}^3$, while a cube constructed in the same fashion is flexible in $\mathbb{R}^3$. The collection of rods and connectors (or framework) shown in Fig. 2, consisting of two triangles with a common edge, is rigid in $\mathbb{R}^2$ but flexible in $\mathbb{R}^3$ since one triangle can then rotate relative to the other along the common edge. (Precise definitions will appear in Section 3.)

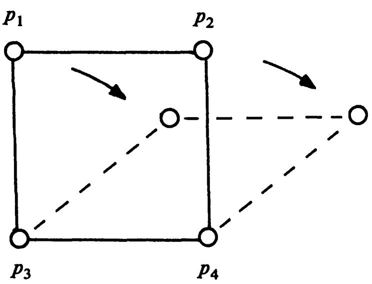


FIG. 1

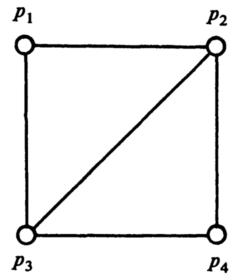


FIG. 2

Is a given framework rigid or flexible in a given Euclidean space $\mathbb{R}^n$? This, the basic question of the subject, does not arrive conveniently labeled as a problem in algebra or analysis or some other field. In fact, contributions have been made to the rigidity problem from a variety of areas, including algebraic geometry, differential topology, complex analysis, projective geometry, linear algebra, graph theory, and combinatorial geometry. Even the Inverse and Implicit Function Theorems from advanced calculus provide powerful tools for the study of rigidity and flexibility, as we shall soon see. The subject simply arises from a real-world problem of some importance and leads to interesting mathematical questions which one attacks by any available means.

The question is hardly new. For well over a century, mathematicians, engineers, and others have studied frameworks under such names as linkages, linkworks, and mechanisms. Their labors produced general results ranging from the theorem of Cauchy **[9]** on the rigidity of surfaces of convex polyhedra to the proof of Kempe **[19]** that any algebraic curve in the plane can be (locally) drawn by an appropriate framework in $\mathbb{R}^2$. But there was widespread interest in specific frameworks as well, ranging from the work of Bricard **[8]**, Bennett **[5]**, and others on flexible octahedra in $\mathbb{R}^3$, to the linkage of Peaucellier described by Hilbert and Cohn-Vossen **[17**, p. 273] which traces out a straight line segment. And attention was also paid to the statics of frameworks, i.e., the resolution of external forces on the framework through the formation of forces of compression and tension in the rods of the framework. A bibliography on linkages compiled by Kanayama **[18]** in 1933, containing over three hundred entries, provides an

The author received his Ph.D. in 1969 under the direction of Reese Prosser. Since then he has been at the University of Wyoming, except for leaves at the University of California, Davis, and at the University of Washington. He was originally interested in rings of continuous functions, ideals of differentiable functions, and distributions. Subsequently, he became fascinated by geodesic domes, and after some practical experience is now writing a book on the rigidity of frameworks.—*Editors*

interesting picture of the variation in activity over the years. The frequent appearance of names such as Cayley, Maxwell, Sylvester, and Tchebychef in the literature indicates the problems were widely known and of broad appeal. Recent years have witnessed a rebirth of interest and some exciting new results in rigidity, ranging from results of Laman **[21]** concerning combinatorial methods in the plane to the work of Bolker and Crapo **[6]**, **[7]** on bracing grids of squares and cubes. Perhaps the most striking modern contribution is the flexible polyhedral surface of Connelly **[10]**, **[11]**, **[12]**, an account of which appears in Kuiper **[20]**. Yet another significant event is the recent appearance of the research bulletin *Structural Topology*, one of whose primary themes is rigidity.

The present paper seeks to impart something of the flavor of the subject by introducing some of its concepts and techniques in an elementary way and then using these to settle one natural and basic question. We begin with a brief look at a few simple examples which may expose the reader to some unfamiliar uses of familiar theorems. After formulating definitions and describing a simple rigidity predictor based on the Inverse Function Theorem, we focus on the following problem. A convex polyhedron C in $\mathbb{R}^3$ gives rise to a framework in a natural way—namely, the framework of vertices and edges of C. Which convex polyhedra give rigid frameworks in $\mathbb{R}^3$? The answer is simply those polyhedra for which every face is a triangle. Its proof relies on ideas which, although originally introduced by Cauchy in 1813, remain among the most beautiful and important in the subject.

Finally, I wish to acknowledge my debt to Herman Gluck—much of the present paper is simply an expanded (and perhaps clarified) version of his paper **[15]**.

2. Elementary Examples.

EXAMPLE 2.1. Consider the square framework shown in Fig. 1, where the vertices initially have coordinates $p_1=(0,1)$, $p_2=(1,1)$, $p_3=(0,0)$, and $p_4=(1,0)$ in $\mathbb{R}^2$. To prevent the square from moving in the plane by translations and rotations (in which the relative positions of the vertices do not change), we fix two vertices of the framework which are joined by an edge, say the third vertex, p_3, and the fourth, p_4. The remaining vertices will be allowed to assume any position consistent with the constraints imposed by the edges.

Let the coordinates of the first and second vertices of the framework be x_1 and x_2, respectively, and let $|\cdot|$ denote the Euclidean norm in $\mathbb{R}^2$. Then the set of solutions of the system of *edge equations*

$$\begin{aligned} &|x_1-x_2|^2=1\\ &|x_1-p_3|^2=|x_1|^2=1\\ &|x_2-p_4|^2=|x_2-(1,0)|^2=1 \end{aligned} \tag{2.1}$$

is the set of possible locations of the first and second vertices of the framework. (In general we allow edges to cross each other and vertices to coincide, ignoring any mechanical problems that might arise in this way.) The family of solutions

$$x_1(t)=(t,\sqrt{1-t^2}\,),\quad x_2(t)=(1+t,\sqrt{1-t^2}\,)\quad \text{for } t\in[0,1]$$

gives the flexing of the square shown in Fig. 1 which begins at $(p_1,p_2)=(0,1,1,1)$, i.e., satisfies $(x_1(0),x_2(0))=(p_1,p_2)$.

EXAMPLE 2.2. We add an edge between the second vertex, p_2, and the third, p_3, of Example 2.1, obtaining the framework shown in Fig. 2. Again fix vertices p_3 and p_4 of the framework and let x_1 and x_2 be the coordinates of the first and second vertices. The system of edge equations now consists of the three equations (2.1) together with the equation

$$|x_2-p_3|^2=|x_2|^2=2.$$

The set of solutions $(x_1,x_2) \in \mathbb{R}^2\times\mathbb{R}^2=\mathbb{R}^4$ of this system of equations is the finite set

$$\{(0,1,1,1),\ (1,0,1,1),\ (0,-1,1,-1),\ (1,0,1,-1)\}$$

as can easily be seen by drawing circles to represent the solution sets of the various edge equations. Since the first and second vertices of the framework cannot continuously move away from their given position $p=(p_1,p_2)=(0,1,1,1)$ while remaining in the solution set of the system of edge equations, the framework is rigid in $\mathbb{R}^2$.

Thus it is the nature of the solution set of the system of edge equations near the given location of the vertices of the framework which determines the rigidity or flexibility of the framework. And to predict the rigidity or flexibility of a framework in the plane, one need *only* solve its system of edge equations, at least near the initial location of its vertices. This very elementary approach to rigidity generalizes to higher dimensional spaces, although some care must be exercised in choosing the vertices to fix in order to eliminate the motions of the framework as a rigid body. For example, in $\mathbb{R}^3$ one can fix three noncollinear vertices, all pairs of which are joined by edges, i.e., fix a triangle.

The Inverse and Implicit Function Theorems provide a somewhat more sophisticated approach to rigidity. It is convenient at this stage to adopt a slightly different point of view and consider the *edge function* of a framework rather than the system of edge equations. The edge function of Example 2.2 is the function f: $\mathbb{R}^4\to\mathbb{R}^4$ defined for $x=(x_1,x_2) \in \mathbb{R}^2\times\mathbb{R}^2=\mathbb{R}^4$ by

$$f(x)=(|x_1-x_2|^2,\ |x_1-p_3|^2,\ |x_2-p_4|^2,\ |x_2-p_3|^2) \tag{2.2}$$

where p_3 and p_4 are the two fixed vertices. The solution set of the system of edge equations is precisely $f^{-1}(f(p))$ where $p=(p_1,p_2)=(0,1,1,1)$.

The Inverse Function Theorem says that a continuously differentiable function f from $\mathbb{R}^n$ to $\mathbb{R}^n$ has a continuously differentiable inverse in a neighborhood of any point $p\in\mathbb{R}^n$ for which the derivative $df(p)$ is a nonsingular linear transformation. Therefore, among other things, the Inverse Function Theorem guarantees that f is one-to-one in a neighborhood of any $p\in\mathbb{R}^n$ for which $df(p)$ has rank n.

EXAMPLE 2.2 (continued). Let f: $\mathbb{R}^4\to\mathbb{R}^4$ be the edge function of the framework in Example 2.2 given in equation (2.2). The matrix with respect to the standard basis for $\mathbb{R}^4$ of the derivative $df(p)$ of f at $p=(p_1,p_2)=(0,1,1,1)$ is obtained by evaluating the partial derivatives of the four coordinate functions of f at p. Rather than writing $df(p)$ as a matrix with four columns, it is very convenient for our purposes to write the matrix $df(p)$ with just two columns where the entries in each column are actually vectors in $\mathbb{R}^2$. Doing this, we find

$$df(p)=2\begin{bmatrix} p_1-p_2 & p_2-p_1 \\ p_1-p_3 & 0 \\ 0 & p_2-p_4 \\ 0 & p_2-p_3 \end{bmatrix}$$

where the first column is obtained from partial derivatives with respect to the coordinates of x_1 and the second column from partial derivatives with respect to the coordinates of x_2. One simple geometrical way to see that $df(p)$ is invertible is to show that its rows are linearly independent. Consider a linear combination of the rows of $df(p)$ which equals zero. Summing over the first column of $df(p)$, we see that the coefficients of the first two rows of $df(p)$ must vanish since p_1-p_2 and p_1-p_3 are clearly linearly independent vectors in $\mathbb{R}^2$. Similarly, summing over the second column of $df(p)$ shows that the coefficients of the last two rows of $df(p)$ also vanish since p_2-p_4 and p_2-p_3 are linearly independent.

By the Inverse Function Theorem, there exists a neighborhood U of $p=(p_1,p_2)$ in $\mathbb{R}^4$ such that f is one-to-one on U. Therefore, $f^{-1}(f(p))\cap U=\{p\}$, i.e., p is the only solution of the system of edge equations in U. The meaning of this is pictured in Fig. 3 where U_1 and U_2 are

neighborhoods of p_1 and p_2, respectively, such that $U_1 \times U_2 \subset U$. The only solution (x_1,x_2) of the edge equations with $x_1 \in U_1$ and $x_2 \in U_2$ is given by $(x_1,x_2)=(p_1,p_2)$, and thus it is not possible to continuously move the first and second vertices of the framework away from their given positions p_1 and p_2 while preserving the edge lengths of the framework. Therefore the framework is rigid in $\mathbb{R}^2$.

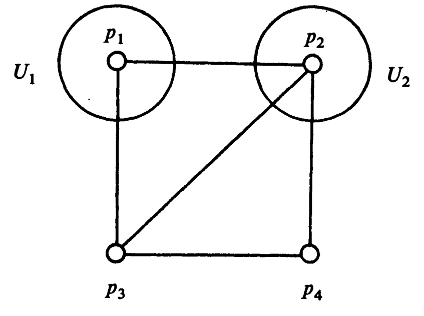


FIG. 3

The reader might find it instructive to verify the rigidity of a tetrahedron in $\mathbb{R}^3$ in the same way, remembering to fix a triangle and considering the matrix $df(p)$ as having just one column with entries that are vectors in $\mathbb{R}^3$.

The Implicit Function Theorem, which can be viewed as providing information about sets of the form $f^{-1}(f(p))$, turns out to be every bit as useful for establishing the flexibility of frameworks as the Inverse Function Theorem for rigidity. Suppose f: $\mathbb{R}^{n+m} \to \mathbb{R}^m$ is continuously differentiable, and let $p=(a,b)\in \mathbb{R}^{n+m}$ where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$. The Implicit Function Theorem says that if the last m columns of $df(p)$ are linearly independent then there exists a neighborhood U of a in $\mathbb{R}^n$ such that there is a unique continuously differentiable function g: $U \to \mathbb{R}^m$ satisfying $g(a)=b$ and

$$(x,g(x)) \in f^{-1}(f(p)) \quad \text{for all } x \in U.$$

EXAMPLE 2.1 (continued). The edge function f: $\mathbb{R}^4 \to \mathbb{R}^3$ of the square framework in Fig. 1 is defined by

$$f(x)=(|x_1-x_2|^2, |x_1-p_3|^2, |x_2-p_4|^2)$$

for

$$x=(x_1,x_2)\in \mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4$$

where p_3 and p_4 are the fixed vertices. It is easy to verify that the last three columns of $df(p)$ are linearly independent where $p=(p_1,p_2)=(0,1,1,1)$. Thus there exists a neighborhood U of 0 in $\mathbb{R}$ and a unique function g: $U \to \mathbb{R}^3$ satisfying $g(0)=(1,1,1)$ and $(t,g(t))\in f^{-1}(f(p))$ for all $t \in U$. Of course, in this particular example, we even know that

$$g(t)=(\sqrt{1-t^2}\,,\ 1+t,\ \sqrt{1-t^2}\,) \quad \text{for all } t \in U \cap [-1,1]$$

since g is unique and

$$(t,\ \sqrt{1-t^2}\,,\ 1+t,\ \sqrt{1-t^2}\,) \in f^{-1}(f(p)) \quad \text{for all } t \in [-1,1].$$

However, the important thing is that $(t,g(t))$ for $t \in U$ gives a flexing of the square satisfying $(0,g(0))=p$.

In general, the Implicit Function Theorem (if applicable) allows one to vary some coordinates of some vertices in a neighborhood U and the function g then prescribes the remaining

coordinates of the vertices in such a way as to remain in the solution set of the system of edge equations. Even if $U \subset \mathbb{R}^n$ where $n>1$, it is then easy to produce a continuous path x beginning at p and lying in the solution set of the system of edge equations, i.e., satisfying $x(0)=p$ and $x(t) \in f^{-1}(f(p))$ for all $t \in [0,1]$. This is essentially the definition of flexibility we adopt in the next section.

3. Definitions. In this section, definitions of frameworks, edge functions, rigidity, and flexibility are formulated in $\mathbb{R}^n$ in order to deal with the most interesting cases ($n=2$ and $n=3$) simultaneously.

An (*abstract*) *framework* G is a set $V=\{1,2,,\ldots,,v\}$ together with a nonempty set E of two-element subsets of V. Each element of V is referred to as a *vertex of* G while each element of E is called an *edge of* G. For $i \in V$, we let $a(i)=\{j \in V: \{i,j\} \in E\}$, the set of vertices of G which are adjacent (or joined by an edge) to the vertex i. Since an (abstract) framework is really nothing other than an (abstract) graph, we occasionally use the language of graph theory. However, our primary interest is not in abstract frameworks but rather in their concrete realizations in some Euclidean space $\mathbb{R}^n$. A *framework* $G(p)$ *in* $\mathbb{R}^n$ is an abstract framework $G=(V,E)$ together with a point

$$p=(p_1,\ldots,p_v) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n = \mathbb{R}^{nv}.$$

We refer to p_i for $i \in V$ as a *vertex of* $G(p)$ and the closed line segment $[p_i,p_j]$ in $\mathbb{R}^n$ for $\{i,j\} \in E$ as an *edge of* $G(p)$. In other words, the framework $G(p)$ in $\mathbb{R}^n$ is obtained by locating vertex i of G at the point $p_i \in \mathbb{R}^n$.

For the remainder of the paper, we dispense with fixing vertices; and thus our definition of the edge function of a framework now takes a slightly different form. Consider a framework $G=(V,E)$ with v vertices and e edges, i.e., $V=\{1,\ldots,v\}$ and E has e elements. Order the e edges of G in some way (lexicographically, if you wish) and define f: $\mathbb{R}^{nv} \to \mathbb{R}^e$, the *edge function of* G, by

$$f(p)=f(p_1,\ldots,p_v)=(\ldots,|p_i-p_j|^2,\ldots)$$

where $\{i,j\} \in E$, $p_k \in \mathbb{R}^n$ for $1 \leqslant k \leqslant v$, and $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^n$. If $G(p)$ is a framework in $\mathbb{R}^n$, then $f(p) \in \mathbb{R}^e$ consists of the squares of the lengths of the e edges of G and thus $f^{-1}(f(p))$ is the set of $q \in \mathbb{R}^{nv}$ such that $G(p)$ and $G(q)$ have corresponding edge lengths equal.

EXAMPLE 3.1. Consider the framework of Example 2.2. shown in Fig. 2. Its abstract framework G is given by $V=\{1,2,3,4\}$ and $E=\{\{1,2\},\{1,3\},\{2,3\},\{2,4\},\{3,4\}\}$ and we examined the framework $G(p)$ in $\mathbb{R}^2$ where

$$p=(p_1,p_2,p_3,p_4)=(0,1,1,1,0,0,1,0).$$

The edge function of G is the map f: $\mathbb{R}^8 \to \mathbb{R}^5$ defined by

$$f(q)=(|q_1-q_2|^2,|q_1-q_3|^2,|q_2-q_3|^2,|q_2-q_4|^2,|q_3-q_4|^2)$$

where

$$q=(q_1,q_2,q_3,q_4) \in \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^8.$$

Unfortunately, the set $f^{-1}(f(p))$ now includes all $q \in \mathbb{R}^{nv}$ that are obtained by simply moving the vertices p of the framework $G(p)$ around by translations, rotations, and, in general, rigid motions of $\mathbb{R}^n$. Recall that a *rigid motion of* $\mathbb{R}^n$ is a distance preserving map T: $\mathbb{R}^n \to \mathbb{R}^n$, i.e., a map satisfying $|Tx-Ty|=|x-y|$ for all x and y in $\mathbb{R}^n$. For $p=(p_1,\ldots,p_v)$ and $q=(q_1,\ldots,q_v)$ in $\mathbb{R}^{nv}$, we say that p and q are *congruent* if there exists a rigid motion T of $\mathbb{R}^n$ such that $Tp_i=q_i$ for $1 \leqslant i \leqslant v$. If $G(p)$ is a framework in $\mathbb{R}^n$ and f is its edge function, then the set $M=\{q \in \mathbb{R}^{nv}: q$ is congruent to $p\}$ is obviously a subset of $f^{-1}(f(p))$ since rigid motions are distance preserving.

The set M of points congruent to p is a smooth manifold (where here and throughout the paper "smooth" means infinitely differentiable). Moreover, if the affine span of the points $p_1,\ldots,p_v$ is $\mathbb{R}^n$ (which means that the points $p_1,\ldots,p_v$ do not lie on any hyperplane in $\mathbb{R}^n$), then M is $n(n+1)/2$-dimensional since it arises from the $n(n-1)/2$-dimensional manifold of orthogonal transformations of $\mathbb{R}^n$ and the n-dimensional manifold of translations of $\mathbb{R}^n$. Thus M is 6-dimensional for $G(p)$ in $\mathbb{R}^3$ and 3-dimensional for $G(p)$ in $\mathbb{R}^2$. We are now in a position to define rigidity and flexibility.

DEFINITION 3.2. Suppose G is a framework with v vertices, f is its edge function, and $p\in\mathbb{R}^{nv}$. The framework $G(p)$ in $\mathbb{R}^n$ is *flexible in* $\mathbb{R}^n$ if there exists a continuous function x: $[0,1]\to\mathbb{R}^{nv}$ satisfying

(i) $x(0)=p$,
(ii) $x(t)\in f^{-1}(f(p))$ for all $t\in[0,1]$, and
(iii) $x(t)$ is not congruent to p for all $t\in(0,1]$.

Such a path x is called a *flexing* of $G(p)$. The framework $G(p)$ is *rigid in* $\mathbb{R}^n$ if it is not flexible in $\mathbb{R}^n$.

Condition (i) says that the path begins at p, (ii) that edge lengths remain constant, and (iii) that for all $t\in(0,1]$, $G(x(t))$ is not obtained by simply moving $G(p)$ as a rigid body. Thus $G(p)$ is flexible in $\mathbb{R}^n$ if and only if the vertices of $G(p)$ can be continuously moved from p to noncongruent positions while preserving the edge lengths of the framework. Note that this definition allows edges to *pass through one another* during a flexing.

The concepts of rigidity and flexibility are invariant under reasonable changes in the definitions. For example, requiring that the path x be infinitely differentiable (or even real analytic) or that $x(t)$ be noncongruent to p for just some $t\in(0,1]$ leads to equivalent notions of flexibility (see Gluck **[15]** or Asimow and Roth **[2]**). However, there is one disconcerting feature of our definition of rigidity. One would like rigidity to mean that every $q\in f^{-1}(f(p))$ sufficiently close to p is actually congruent to p (which is the analog of the behavior shown in Fig. 3 in the present setting with no vertices fixed) and it is far from clear that our definition guarantees this. There might exist points in $f^{-1}(f(p))$ arbitrarily close to p but not in M while no flexing of $G(p)$ exists. However, general results in algebraic geometry regarding the existence of paths in algebraic sets (see Milnor **[22**, Lemma 3.1]) prevent this occurrence, and it is indeed the case that $G(p)$ is rigid in $\mathbb{R}^n$ if and only if M and $f^{-1}(f(p))$ coincide near p.

4. Infinitesimal Flexibility and Statics. We now introduce the concept of infinitesimal flexibility which arises from that of flexibility by focusing on the tangential conditions imposed by the preservation of edge length requirements. Suppose G is a framework with v vertices, f is its edge function, and $p\in\mathbb{R}^{nv}$. Let $x=(x_1,\ldots,x_v)$ be a smooth function on $[0,1]$ satisfying $x(0)=p$ and $x(t)\in f^{-1}(f(p))$ for all $t\in[0,1]$. Thus for all edges $\{i,j\}$ of G, we have

$$|x_i(t)-x_j(t)|^2=|p_i-p_j|^2 \quad \text{for all } t\in[0,1].$$

Differentiating and evaluating at $t=0$, one obtains

$$(x_i(0)-x_j(0))\cdot(x_i'(0)-x_j'(0))=(p_i-p_j)\cdot(x_i'(0)-x_j'(0))=0$$

for all edges $\{i,j\}$ of G. Thus a smooth flexing x of $G(p)$ assigns a velocity vector $\mu_i=x_i'(0)\in\mathbb{R}^n$ to each vertex p_i of $G(p)$ in such a way that

$$(p_i-p_j)\cdot(\mu_i-\mu_j)=0 \text{ for all edges } \{i,j\} \text{ of } G. \tag{4.1}$$

In terms of the edge function, it is easy to see that $\mu=(\mu_1,\ldots,\mu_v)\in\mathbb{R}^{nv}$ satisfies (4.1) if and only if μ is an element of the kernel of the linear transformation $df(p)$: $\mathbb{R}^{nv}\to\mathbb{R}^e$. For by applying the matrix of $df(p)$ to μ, one finds $df(p)(\mu)=0$ if and only if

$$(p_i-p_j)\cdot\mu_i+(p_j-p_i)\cdot\mu_j=(p_i-p_j)\cdot(\mu_i-\mu_j)=0$$

for all edges $\{i,j\}$ of G.

Recall now that a flexing $x(t)$ of $G(p)$ begins at p, preserves edge lengths, and also does not belong to the manifold M of points congruent to p for all $t>0$. In the same spirit, we require that an infinitesimal flexing μ of $G(p)$ instantaneously preserve edge lengths, i.e., satisfy (4.1), and also not belong to the tangent space T_p to the manifold M at the point p. Note that $T_p \subset$ kernel $df(p)$, since if $x=(x_1,\ldots,x_v)$: $\mathbb{R}\to M$ is a smooth path with $x(0)=p$ then for all i and j we have

$$|x_i(t)-x_j(t)|^2=|p_i-p_j|^2 \quad \text{for all } t\in\mathbb{R}$$

and hence $\mu = x'(0)$ satisfies (4.1). These observations lead to the following definition.

DEFINITION 4.1. Suppose G is a framework with v vertices, f is its edge function, and $p\in\mathbb{R}^{nv}$. The framework $G(p)$ is *infinitesimally rigid in* $\mathbb{R}^n$ if $T_p=$kernel $df(p)$ and *infinitesimally flexible in* $\mathbb{R}^n$ otherwise. Elements of kernel $df(p)-T_p$ are called *infinitesimal flexings of* $G(p)$.

How are infinitesimal flexibility and flexibility related? In Section 5, we show that flexibility implies infinitesimal flexibility, and the following example establishes that infinitesimal flexibility is a strictly weaker notion.

EXAMPLE 4.2. Consider the degenerate triangle $G(p)$ in $\mathbb{R}^2$ shown in Fig. 4 with collinear vertices, say $p_1=(0,0)$, $p_2=(1,0)$, and $p_3=(2,0)$. Then $G(p)$ is infinitesimally flexible but not flexible in $\mathbb{R}^2$. For if we let $\mu_1=\mu_3=0$ and $\mu_2=(0,c)$ where $c\neq 0$, then (4.1) is satisfied but $\mu=(\mu_1,\mu_2,\mu_3)$ is clearly not the derivative at $t=0$ of a smooth motion of $G(p)$ as a rigid body in $\mathbb{R}^2$. However, it is easy to see that no flexing of $G(p)$ exists.

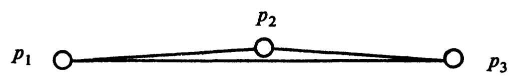


FIG. 4

Historically, the existence of these two closely related but distinct notions of rigidity and the absence of adequate definitions of both have combined to create widespread confusion. Even today, many engineering textbooks contain material that is, at best, misleading and, at worst, just plain false in connection with rigidity. Grünbaum and Shephard **[16]** provide a vivid account of this unfortunate situation. The evidence strongly suggests that engineers are primarily interested in infinitesimal rigidity rather than rigidity. This preference is explained by the close connections between infinitesimal rigidity and the study of statics for frameworks (which have led some authors to refer to infinitesimal rigidity as static rigidity). We now embark on a very short course on statics in $\mathbb{R}^3$ where the reader should think of frameworks in $\mathbb{R}^3$ as actual physical objects whose edges are straight, stiff rods that are connected by articulated joints at the vertices.

Certain systems of internal forces in a framework provide a convenient beginning for this discussion. Consider a framework $G(p)$ in $\mathbb{R}^3$ where $p\in\mathbb{R}^{3v}$ and suppose the various rods of the framework are subject to forces of compression or tension directed along the rods. More precisely, suppose there is associated with each rod $[p_i,p_j]$ a scalar $\omega_{\{i,j\}}$ such that $\omega_{\{i,j\}}(p_i-p_j)$ is the force exerted by the rod on the vertex p_i and $\omega_{\{i,j\}}(p_j-p_i)$ is the force exerted by the rod on the vertex p_j. The scalar $\omega_{\{i,j\}}$ thus gives the magnitude of the force *per unit length.* If $\omega_{\{i,j\}}<0$, the force is called a *tension* in the rod; while if $\omega_{\{i,j\}}>0$, the force is referred to as a *compression.* Note that the rod exerts forces on the vertices p_i and p_j which are equal in magnitude but opposite in direction. In general, a vertex of $G(p)$ will be incident with several rods and our interest focuses on the situation in which the sum of the forces exerted on each vertex equals zero. A *stress* of a framework $G(p)$ in $\mathbb{R}^3$ is a collection of scalars $\omega_{\{i,j\}}$, one for each edge $[p_i,p_j]$ of $G(p)$, such that

$$\sum_{j\in a(i)} \omega_{\{i,j\}}(p_i-p_j)=0 \quad \text{for } 1\leqslant i\leqslant v.$$

Letting $\omega_{\{i,j\}}=0$ for all edges gives the trivial stress, and we say a framework is *stress free* if it admits only the trivial stress. The degenerate triangle $G(p)$ of Example 4.2 shown in Fig. 4 has a nontrivial stress given by $\omega_{\{1,2\}}=\omega_{\{2,3\}}=-2$ and $\omega_{\{1,3\}}=1$. One can imagine this stress arising from an attempt to construct the degenerate triangle with rod $[p_1,p_3]$ slightly too long, creating a compression in this rod and tensions in the other two rods.

Stresses have a convenient description in terms of the edge function of a framework. Since the edge function $f:\mathbb{R}^{3v}\to\mathbb{R}^e$ of a framework G with v vertices and e edges is defined for $q\in\mathbb{R}^{3v}$ by

$$f(q)=(\ldots,|q_i-q_j|^2,\ldots)$$

where $\{i,j\}$ is an edge of G, we see that the matrix of $df(p)$ has $3v$ columns, each triple of columns arising from the partial derivatives with respect to the three coordinates of a vertex p_i of $G(p)$. And the matrix of $df(p)$ has e rows, each arising from an edge $\{i,j\}$ of G. The $\{i,j\}$ row of $df(p)$, obtained by taking partial derivatives of $|q_i-q_j|^2$ and evaluating at p, has only two triples of nonzero entries, namely, $2(p_i-p_j)$ in the column triple corresponding to p_i and $2(p_j-p_i)$ in the column triple corresponding to p_j. Thus the $e\times 3v$ matrix of $df(p)$ has the form

$$\begin{array}{cc} & \begin{array}{ccccccc} \text{vertex } 1 & \cdots & \text{vertex } i & \cdots & \text{vertex } j & \cdots & \text{vertex } v \end{array} \\ \text{edge } \{i,j\} & \left[\begin{array}{ccccccc} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 2(p_i-p_j) & \cdots & 2(p_j-p_i) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array}\right] \end{array}.$$

In light of this, it is clear that a stress of a framework $G(p)$ in $\mathbb{R}^3$ is nothing other than the collection of coefficients of a linear dependence among the rows of $df(p)$ (or, equivalently, an element of the kernel of $df(p)^t$: $\mathbb{R}^e\to\mathbb{R}^{3v}$, the transpose of $df(p)$). Thus a framework $G(p)$ in $\mathbb{R}^3$ is stress free if and only if rank $df(p)=e$, the number of edges of the framework.

The next stage in our study of statics allows external forces to act on the framework. If a set of external forces is applied to the vertices of a framework, one force for each vertex, then forces of tension and compression presumably arise in the rods of the framework. Can the framework resolve the set of external forces in the sense that at each vertex the sum of all the forces, internal and external, is zero? If so, is the resolution unique?

Consider a framework $G(p)$ in $\mathbb{R}^3$. Since our interest is in statics, we concentrate on external forces of the following type. A vector $F=(F_1,\ldots,F_v)\in\mathbb{R}^{3v}$ is an *equilibrium force for* $p=(p_1,\ldots,p_v)$ if

$$\sum_{i=1}^{v} F_i=0 \quad\text{and}\quad \sum_{i=1}^{v} p_i\times F_i=0$$

where $\times$ denotes the cross product in $\mathbb{R}^3$. The first condition merely says that the sum of the forces F_i is zero while the second condition says that the sum of the moments (or torques) $p_i\times F_i$ about any axis through the origin of the forces F_i applied at the points p_i is zero. Together, the conditions imply that the sum of the moments about any axis is zero. On the other hand, we say that $F=(F_1,\ldots,F_v)\in\mathbb{R}^{3v}$ is a *resolvable force for* $G(p)$ if there exist scalars $\omega_{\{i,j\}}$, one for each edge $[p_i,p_j]$ of $G(p)$, such that

$$F_i+\sum_{j\in a(i)}\omega_{\{i,j\}}(p_i-p_j)=0 \quad \text{for } 1\leqslant i\leqslant v.$$

This condition means that the sum of all the forces at every vertex is zero. In these terms, a stress of $G(p)$ is a resolution of the trivial force $F=(0,\ldots,0)$ for p.

Let $\mathfrak{E}$ and $\mathfrak{R}$ be the collections of all equilibrium and resolvable forces for $G(p)$, respectively. Both $\mathfrak{E}$ and $\mathfrak{R}$ are subspaces of $\mathbb{R}^{3v}$ since $\mathfrak{E}$ is the kernel of the linear map L: $\mathbb{R}^{3v} \to \mathbb{R}^6$ defined by

$$L(F_1,\ldots,F_v)=(\Sigma F_i, \Sigma p_i \times F_i)$$

and $\mathfrak{R}$ is the image of the transpose $df(p)^t$: $\mathbb{R}^e \to \mathbb{R}^{3v}$ of $df(p)$. It is easy to show that $\mathfrak{R} \subset \mathfrak{E}$ since $F=(F_1,\ldots,F_v) \in \mathfrak{R}$ means that F is a linear combination of the rows of $df(p)$ and each row of $df(p)$ satisfies the conditions defining $\mathfrak{E}$. That is, p_i-p_j and p_j-p_i are the only nonzero entries in the $\{i, j\}$ row of $df(p)$; so we have $\Sigma F_i=0$. And $\Sigma p_i \times F_i=0$ since $p_i \times (p_i-p_j)+p_j \times (p_j-p_i)=0$.

The dimension of the vector space $\mathfrak{R}$ of resolvable forces for $G(p)$ is rank $df(p)$ since $\mathfrak{R}=$ image $df(p)^t$. If $p_1,\ldots,p_v$ are not coplanar in $\mathbb{R}^3$, then $\mathfrak{E}$ has dimension $3v-6$ since L maps $\mathbb{R}^{3v}$ onto $\mathbb{R}^6$ in this case. To see this, suppose p_2-p_1, p_3-p_1, and p_4-p_1 span $\mathbb{R}^3$ and consider arbitrary vectors x and y in $\mathbb{R}^3$. There exist scalars λ_1, λ_2, and λ_3 such that

$$x=\sum_{i=1}^{3} \lambda_i(p_{i+1}-p_1).$$

Let $F=(F_1,\ldots,F_v)=(-\Sigma\lambda_i y, \lambda_1 y, \lambda_2 y, \lambda_3 y, 0,\ldots,0)$. Then it is easy to verify that

$$L(F)=(\Sigma F_i, \Sigma p_i \times F_i)=(0, x \times y).$$

Therefore image L contains all vectors of the form $(0,z)$ for $z \in \mathbb{R}^3$. Since image L also contains all vectors of the form $(z, p_1 \times z)$ for $z \in \mathbb{R}^3$, L is clearly onto.

We are now in a position to relate infinitesimal rigidity to the resolvability of equilibrium forces.

PROPOSITION 4.3. *Suppose $G(p)$, $p=(p_1,\ldots,p_v) \in \mathbb{R}^{3v}$, is a framework in $\mathbb{R}^3$ where $p_1,\ldots,p_v$ are not coplanar. Then $G(p)$ is infinitesimally rigid in $\mathbb{R}^3$ if and only if every equilibrium force for p is a resolvable force for $G(p)$. Moreover, each equilibrium force is uniquely resolvable if and only if $G(p)$ is infinitesimally rigid and stress free.*

Proof. Since dimension $\mathfrak{R}=$ rank $df(p)$, dimension $\mathfrak{E}=3v-6$, and $\mathfrak{R} \subset \mathfrak{E}$, we have $\mathfrak{E} \subset \mathfrak{R}$ if and only if rank $df(p)=3v-6$. But the tangent space T_p to the manifold M is 6-dimensional, since the points $p_1,\ldots,p_v$ are not coplanar, and thus $G(p)$ is infinitesimally rigid in $\mathbb{R}^3$ if and only if dimension kernel $df(p)=6$, which is equivalent to rank $df(p)=3v-6$. The uniqueness result follows without difficulty from the fact that $G(p)$ is stress free if and only if kernel $df(p)^t$ is trivial, which is equivalent to the unique resolvability of the trivial equilibrium force.

5. A Rigidity Predictor. Definition 3.2 and the subsequent discussion make clear that the rigidity or flexibility of a framework $G(p)$ in $\mathbb{R}^n$ is determined by the nature of the inclusion near p of the manifold M of points congruent to p in the algebraic set $f^{-1}(f(p))$. If $f^{-1}(f(p))$ happens to be a manifold of known dimension near p, then the rigidity or flexibility of $G(p)$ is governed by the dimensions of the two manifolds. This is precisely the situation we direct our attention to here.

For a smooth map f: $\mathbb{R}^n \to \mathbb{R}^m$, let $k=\max\{\text{rank } df(x): x \in \mathbb{R}^n\}$, the maximum of the rank of the derivative of f. We say that $p \in \mathbb{R}^n$ is a *regular point of f* if rank $df(p)=k$. The Implicit Function Theorem implies that $f^{-1}(f(p))$ is an $(n-k)$-dimensional smooth manifold near p provided p is a regular point of f (see Auslander and MacKenzie [**4**, Implicit-Parametrization Theorem, p. 32]).

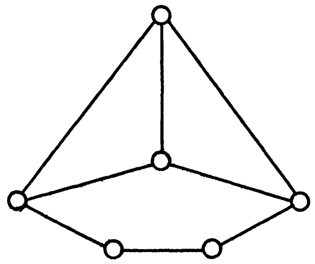
Consider a framework $G(p)$, $p=(p_1,\ldots,p_v) \in \mathbb{R}^{nv}$, in $\mathbb{R}^n$ with edge function f: $\mathbb{R}^{nv} \to \mathbb{R}^e$. Suppose $p_1,\ldots,p_v$ do not lie on a hyperplane in $\mathbb{R}^n$ and let M be the $n(n+1)/2$-dimensional manifold of points congruent to p. Since the tangent space T_p to M at p is a subset of kernel $df(p)$ by the comment preceding Definition 4.1, we have

$$\text{rank } df(p)=nv-\text{dimension kernel } df(p) \leqslant nv-n(n+1)/2. \tag{5.1}$$

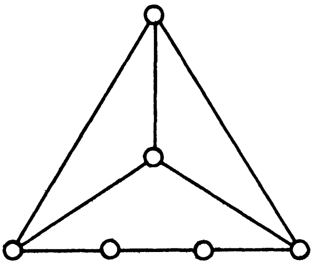
If p is a regular point of f where rank $df(p)=k$, then much more can be said. For in this case, $f^{-1}(f(p))$ is a $(nv-k)$-dimensional manifold near p, and thus M and $f^{-1}(f(p))$ agree near p if and only if their dimensions are equal. Since $G(p)$ is rigid in $\mathbb{R}^n$ if and only if M and $f^{-1}(f(p))$ coincide near p, we conclude that $G(p)$ is rigid in $\mathbb{R}^n$ if and only if rank $df(p)=nv-n(n+1)/2$. This gives the following rigidity predictor which was introduced by Gluck **[15]** and extensively used in Asimow and Roth **[2]**.

PROPOSITION 5.1. *Let $G(p)$ be a framework in $\mathbb{R}^n$ where $p=(p_1,\ldots,p_v)\in\mathbb{R}^{nv}$ is a regular point of the edge function f and $p_1,\ldots,p_v$ do not lie on a hyperplane in $\mathbb{R}^n$. Then $G(p)$ is rigid in $\mathbb{R}^n$ if and only if rank $df(p)=nv-n(n+1)/2$ and $G(p)$ is flexible in $\mathbb{R}^n$ if and only if rank $df(p)$ $<nv-n(n+1)/2$.*

One application of the rigidity predictor leads to the notion of the "generic" behavior of an abstract framework in $\mathbb{R}^n$. Can rigidity in $\mathbb{R}^n$ be considered a property of an abstract framework G rather than just a property of particular realizations of the framework in $\mathbb{R}^n$? The two frameworks in the plane shown in Fig. 5 are given by the same abstract framework, but the



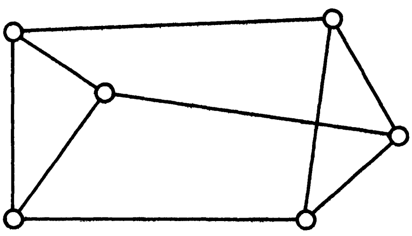
(a)



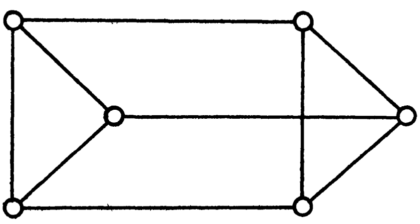
(b)

FIG. 5

framework in (a) is flexible in $\mathbb{R}^2$ (its bottom edges flop around) while the framework in (b) is rigid in $\mathbb{R}^2$ (due to the collinearity of its bottom edges). Therefore, rigidity is not determined solely by the abstract structure of the framework; the location of the vertices must also be taken into account. Another example appears in Fig. 6 where the framework in (a) is rigid in $\mathbb{R}^2$ while



(a)



(b)

FIG. 6

the framework in (b) is flexible in $\mathbb{R}^2$, even though their underlying abstract frameworks are the same. However, the reader may have noticed that the vertex location in (b) in each case is rather carefully contrived while, in some sense, (a) in each case represents the typical behavior of the framework in $\mathbb{R}^2$.

To understand this phenomenon, it is useful to observe that the set of regular points of f is a dense open subset of $\mathbb{R}^{nv}$. For x is a regular point of f if and only if $P(x)\neq 0$, where $P(x)$ is the polynomial given by the sum of the squares of the determinants of all $k\times k$ submatrices of $df(x)$. Since the rigidity predictor says that the regular realizations p (with affine span $\mathbb{R}^n$) of a given

framework are either all rigid in $\mathbb{R}^n$ or all flexible in $\mathbb{R}^n$, we see that every abstract framework has a typical or generic behavior in $\mathbb{R}^n$. In other words, given an abstract framework G, we have either $G(p)$ rigid in $\mathbb{R}^n$ for a dense open set of $p\in\mathbb{R}^{nv}$ or $G(p)$ flexible in $\mathbb{R}^n$ for a dense open set of $p\in\mathbb{R}^{nv}$.

The rigidity predictor and the notion of a regular point also serve to clarify the relationship between rigidity and infinitesimal rigidity. Consider a framework $G(p)$ in $\mathbb{R}^n$ where $p\in\mathbb{R}^{nv}$ is a regular point of the edge function f. In this case, $f^{-1}(f(p))$ is a manifold near p whose tangent space at p is kernel $df(p)$. Since M and $f^{-1}(f(p))$ agree near p if and only if their tangent spaces T_p and kernel $df(p)$ at p are equal and the latter is precisely the definition of infinitesimal rigidity, we conclude that $G(p)$ is rigid in $\mathbb{R}^n$ if and only if $G(p)$ is infinitesimally rigid in $\mathbb{R}^n$. Thus at regular points, rigidity (flexibility) and infinitesimal rigidity (infinitesimal flexibility) are equivalent. Moreover, the following proposition shows that infinitesimal rigidity occurs only at regular points.

PROPOSITION 5.2. *Suppose $G(p)$, $p=(p_1,\ldots,p_v)\in\mathbb{R}^{nv}$, is a framework in $\mathbb{R}^n$ and the affine span of $p_1,\ldots,p_v$ is $\mathbb{R}^n$. Then $G(p)$ is infinitesimally rigid in $\mathbb{R}^n$ if and only if p is a regular point of f and $G(p)$ is rigid in $\mathbb{R}^n$.*

Proof. In light of the above observations, it suffices to show that if $G(p)$ is infinitesimally rigid in $\mathbb{R}^n$, then p is a regular point. The intersection of two dense open sets gives a regular point $q=(q_1,\ldots,q_v)\in\mathbb{R}^{nv}$ such that $q_1,\ldots,q_v$ do not lie on a hyperplane in $\mathbb{R}^n$. If $G(p)$ is infinitesimally rigid in $\mathbb{R}^n$, then $T_p=$kernel $df(p)$ which gives $n(n+1)/2=nv-\text{rank}\,df(p)$. By inequality (5.1) and the fact that q is a point of maximum rank, one obtains

$$\text{rank } df(q)\leqslant nv-n(n+1)/2=\text{rank } df(p)\leqslant \text{rank } df(q)$$

which says that p is a regular point of f.

6. Frameworks Given by Convex Polyhedra in $\mathbb{R}^3$. Let C be a convex polyhedron in $\mathbb{R}^3$, i.e., the convex hull of a finite set of noncoplanar points in $\mathbb{R}^3$. A vertex of C is a point which is the intersection of C with a support plane of C, while an edge of C is a closed line segment which is the intersection of C with a support plane of C. Suppose C has v vertices with coordinates $p_1,\ldots,p_v\in\mathbb{R}^3$. Let $V=\{1,\ldots,v\}$ and

$$E=\{\{i,j\}:[p_i,p_j]\text{ is an edge of } C\}.$$

Then we refer to $G(p)$ where $G=(V,E)$ and $p=(p_1,\ldots,p_v)$ as the framework in $\mathbb{R}^3$ *given by C*. Which convex polyhedra in $\mathbb{R}^3$ give rigid frameworks in $\mathbb{R}^3$ and which give flexible frameworks in $\mathbb{R}^3$?

Suppose $G(p)$ is the framework given by a convex polyhedron C in $\mathbb{R}^3$. We now show that $G(p)$ is stress free, i.e., rank $df(p)=e$ where $f:\mathbb{R}^{3v}\to\mathbb{R}^e$ is the edge function of G and e is the number of edges of G. This implies that $\max\{\text{rank } df(x): x\in\mathbb{R}^{3v}\}=e$ and thus p is a regular point of f. Consequently, the rigidity predictor (Proposition 5.1) with $n=3$ allows one to determine the rigidity or flexibility of $G(p)$ by the simplest imaginable procedure—just count the number of edges of C and compare the result to $3v-6$. Furthermore, the results can even be interpreted "infinitesimally" if desired, since rigidity (flexibility) and infinitesimal rigidity (infinitesimal flexibility) amount to the same thing at regular points.

The proof that rank $df(p)=e$ has two parts, both of which originate in Cauchy's proof **[9]** of the fact that two convex polyhedra in $\mathbb{R}^3$ with corresponding faces congruent and "arranged in the same way" are themselves congruent. (More formally, the hypothesis of Cauchy's Theorem says that there exists a one-to-one correspondence ψ between the sets of vertices of the two polyhedras such that S is the set of vertices of a face of one polyhedron if and only if $\psi(S)$ is the set of vertices of a face of the other and, furthermore, the map ψ preserves distances between vertices on corresponding faces.) One part is of a topological nature and deals with graphs on a

polyhedron, while the other is of a geometrical nature and relies on the convexity of the polyhedron. The particular arrangement of ideas used here is due to Alexandrov **[1]** and Gluck **[15]**, and related results appear in Dehn **[14]** and Weyl **[23]**.

Consider a framework $G(p)$ in $\mathbb{R}^3$ given by a convex polyhedron C and suppose there exists a nontrivial linear combination of the rows of $df(p)$ which vanishes; say $\omega_{\{i,j\}}$ denotes the coefficient of the $\{i,j\}$ row in this linear combination. Summing each column triple of $df(p)$, we find that

$$\sum_{j \in a(i)} \omega_{\{i,j\}}(p_i - p_j) = 0 \quad \text{for } 1 \leqslant i \leqslant v \tag{6.1}$$

where $a(i) = \{j : [p_i, p_j]$ is an edge of $G(p)\}$, the set of vertices adjacent to vertex i.

The signs of the coefficients $\omega_{\{i,j\}}$ are now used to attach the symbols $+$ and $-$ to some of the edges of C. If $\omega_{\{i,j\}} > 0$, then the $\{i,j\}$ edge of C is marked $+$ while if $\omega_{\{i,j\}} < 0$, then the $\{i,j\}$ edge of C is marked $-$. The edge $\{i,j\}$ is left unmarked if $\omega_{\{i,j\}} = 0$. Consider the graph G' on the surface ∂C of C induced by the marked edges of C, which means that the edges of G' are the edges of C marked $+$ or $-$ and the vertices of G' are the vertices of C incident with at least one edge marked $+$ or $-$. For each vertex p_i of G', the edges of G' incident with p_i can be cyclically ordered according to their occurrence on the surface of C as the vertex p_i is circled once (say, in a counterclockwise direction with respect to an outward normal of ∂C). The *index of* p_i is the number of changes of sign encountered in this cycle of edges around the vertex p_i and the *index* I is the sum of the indices of the vertices of G'. The topological part of the proof deals with graphs induced by some subset of the edges of a convex polyhedron where each edge in the subset is marked $+$ or $-$.

LEMMA 6.1. *The index satisfies*

$$I \leqslant 4v' - 8$$

where v' is the number of vertices of G'.

Proof. Let e' be the number of edges of G' and f' the number of regions (or topological components) of $\partial C - G'$. Each such region has a boundary consisting of edges of G'. Let f_n' be the number of regions with exactly n boundary edges where an edge is counted twice for a region if the region lies on both sides of the edge. Clearly $f_1' = 0$ and f_2' is nonzero only when G' has just one edge. Since $I = 0 = 4v' - 8$ in this case, we assume $f_2' = 0$. Then

$$2e' = \sum_{n \geqslant 3} n f_n' \quad \text{and} \quad f' = \sum_{n \geqslant 3} f_n'.$$

We now compute the index I by circling regions rather than vertices. Since the number of sign-changes as one traverses the boundary of a region with n edges is an even number less than or equal to n, we have

$$I \leqslant 2f_3' + 4f_4' + 4f_5' + 6f_6' + 6f_7' + \cdots \leqslant \sum_{n \geqslant 3} (2n-4) f_n'$$

$$= 2 \sum_{n \geqslant 3} n f_n' - 4 \sum_{n \geqslant 3} f_n' = 4e' - 4f'.$$

By Euler's formula, $v' - e' + f' = 1 + N \geqslant 2$ where N is the number of components of the graph G'. Therefore

$$I \leqslant 4e' - 4f' \leqslant 4v' - 8.$$

Next, we present the geometrical part of the argument which relies on the convexity of C together with equation (6.1).

LEMMA 6.2. *The index of every vertex of G' is greater than or equal to four.*

Proof. Consider any vertex p_i of G' and let $a'(i)=\{j:[p_i,p_j]$ is an edge of $G'\}$. By equation (6.1), we have

$$\sum_{j\in a'(i)} \omega_{\{i,j\}}(p_i-p_j)=0 \tag{6.2}$$

since the coefficients $\omega_{\{i,j\}}$ of edges of G but not G' are zero. First, the index of p_i cannot be zero since the scalars $\omega_{\{i,j\}}$ for $j\in a'(i)$ are either all positive or all negative in this case. By the convexity of C, there exists a plane in $\mathbb{R}^3$ which intersects C only at p_i; say an equation of the plane is $n\cdot(p_i-x)=0$, where $n\in\mathbb{R}^3$ is a normal to the plane. Since all the vertices of G' except p_i lie on one side of the plane, $n\cdot(p_i-p_j)$ is either positive for all $j\in a'(i)$ or negative for all $j\in a'(i)$. Therefore,

$$\sum_{j\in a'(i)} \omega_{\{i,j\}}[n\cdot(p_i-p_j)]\neq 0,$$

which is impossible since (6.2) gives

$$\sum_{j\in a'(i)} \omega_{\{i,j\}}[n\cdot(p_i-p_j)]=n\cdot\left[\sum_{j\in a'(i)} \omega_{\{i,j\}}(p_i-p_j)\right]=0.$$

Moreover, a similar argument shows that the index of p_i cannot be two, since in this case there is a set of edges of G' marked $+$ followed by a set of edges marked $-$ in the cycle of edges around p_i. By the convexity of C, there exists a plane through p_i with the edges of G' incident with p_i marked $+$ on one side of the plane and those marked $-$ on the other side of the plane. If an equation of this plane is $n\cdot(p_i-x)=0$, we have $n\cdot(p_i-p_j)$ of one sign for all the edges marked $+$ and of the opposite sign for those marked $-$. Thus by (6.2)

$$0=n\cdot\left[\sum_{j\in a'(i)} \omega_{\{i,j\}}(p_i-p_j)\right]=\sum_{j\in a'(i)} \omega_{\{i,j\}}[n\cdot(p_i-p_j)]\neq 0.$$

This contradiction completes the proof of Lemma 6.2.

THEOREM 6.3. *Let $G(p)$, $p\in\mathbb{R}^{3v}$, be the framework in $\mathbb{R}^3$ given by a convex polyhedron C and suppose f is the edge function of G. Then*

$$\operatorname{rank} df(p)=e,$$

the number of edges of C.

Proof. We suppose that $G(p)$ admits a nontrivial stress and arrive at a contradiction. We attach the symbols $+$ and $-$ to some of the edges of C according to the signs in a nontrivial stress and let G' be the graph induced by the marked edges of C. By Lemmas 6.1 and 6.2 we have

$$I\leqslant 4v'-8<4v'\leqslant I$$

where I is the index and v' the number of vertices of G'. This contradiction shows that $G(p)$ is stress free and thus rank $df(p)=e$.

Therefore, in light of the rigidity predictor, the rigidity or flexibility of a framework arising from a convex polyhedron is determined by a simple comparison of e and $3v-6$. However, this same comparison arises in another quite different way. Consider a convex polyhedron C in $\mathbb{R}^3$ with v vertices, e edges, and f faces of which f_n have exactly n edges. By Euler's formula,

$$3v-6=3(v-2)=3(e-f)=e+(2e-3f).$$

But

$$3f=3\sum_{n\geqslant 3} f_n\leqslant \sum_{n\geqslant 3} nf_n=2e$$

with equality if and only if $f=f_3$, i.e., every face of C is a triangle. Therefore $e\leqslant 3v-6$ with equality if and only if every face of C is a triangle, which leads to the following corollary.

COROLLARY 6.4. *The framework $G(p)$ given by a convex polyhedron C is rigid in $\mathbb{R}^3$ if and only if every face of C is a triangle.*

Proof. by Theorem 6.3, rank $df(p)=e$ where e is the number of edges of C and f is the edge function of G. Therefore, $p=(p_1,\dots,p_v)$ is a regular point of f and clearly $p_1,\dots,p_v$ are not coplanar. By the rigidity predictor, $G(p)$ is rigid in $\mathbb{R}^3$ if and only if $e=\text{rank } df(p)=3v-6$. But, as we just observed, $e=3v-6$ if and only if every face of C is a triangle.

7. Concluding Remarks. The $e=3v-6$ test for rigidity arising from Theorem 6.3 and the rigidity predictor has certainly not escaped the attention of engineers. In fact, the inaccuracies that mar many accounts of rigidity stem from attempts to apply this simple formula to all frameworks in $\mathbb{R}^3$. It is not difficult to find examples showing this is inappropriate.

For instance, consider the framework $G(p)$ in $\mathbb{R}^3$ shown in Fig. 7, which is a tetrahedron with a triangle in the plane of its base. A simple geometrical argument using the very special location of the three vertices in the interior of the base of the tetrahedron shows that $G(p)$ is rigid in $\mathbb{R}^3$ even though $e<3v-6$. (Note Theorem 6.3 is not applicable since $G(p)$ is not the framework given by a convex polyhedron in $\mathbb{R}^3$ in the sense defined in Section 6.) However, it is quite easy to show that the framework G (in fact, any framework with $e<3v-6$) is generically flexible or, equivalently, always infinitesimally flexible in $\mathbb{R}^3$.

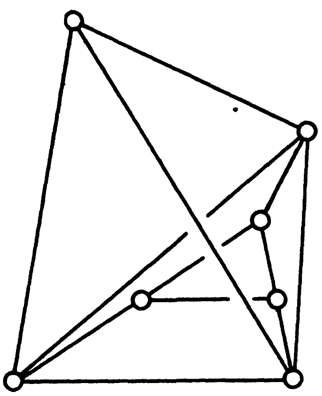


FIG. 7

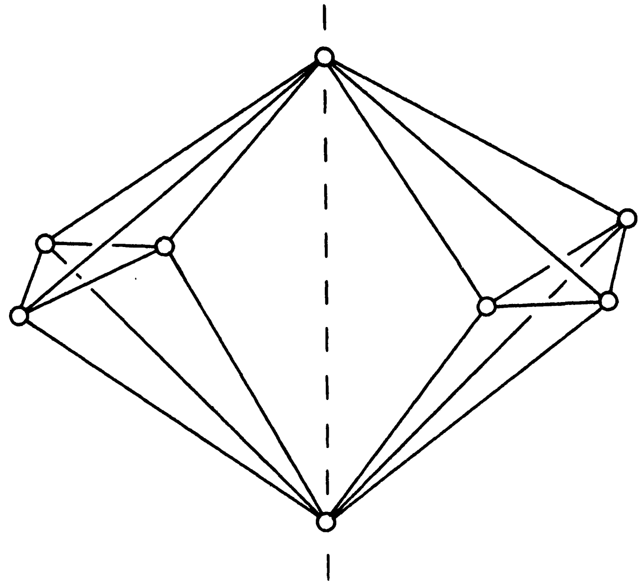


FIG. 8

On the other hand, there exist frameworks with $e=3v-6$ which are not only flexible but even generically flexible in $\mathbb{R}^3$. For example, the framework $G(p)$ shown in Fig. 8 is flexible in $\mathbb{R}^3$ since one half of the framework rotates relative to the other around the dotted line shown, and this is clearly the typical behavior of G in $\mathbb{R}^3$. However, for planar graphs with at least three vertices, it can be shown using Steinitz's Theorem that $e=3v-6$ is a necessary and sufficient condition for generic rigidity in $\mathbb{R}^3$. Of course, this does not preclude the existence of (a small set of) flexible realizations of a planar graph satisfying $e=3v-6$. For instance, Bricard **[8]** describes all of the flexible realizations of the framework arising from an octahedron in $\mathbb{R}^3$. Since the surfaces of all these flexible octahedra are self-intersecting, one can conclude that the framework given by any embedded octahedron is rigid in $\mathbb{R}^3$. The conjecture, a version of which Euler proposed in 1766, that all embedded polyhedral surfaces are rigid, has recently been settled by

an ingenious counterexample of Connelly **[10]**, **[11]**, **[12]**. This flexible polyhedral surface is a collection of triangles joined together along common edges in such a way as to form a closed (but not convex) polyhedron without self-intersections in $\mathbb{R}^3$. Its remarkable property is that the collection of vertices and edges of the triangles forms a flexible framework in $\mathbb{R}^3$.

While Corollary 6.4 tells us that the framework given by a cube is flexible in $\mathbb{R}^3$, it gives no information about the "braced" cube obtained by adding new diagonal edges across some (or perhaps all) of the six faces of the cube. Since such frameworks are surely of interest, we conclude with a short discussion of frameworks which arise in various more general ways from convex polyhedra in $\mathbb{R}^3$. Incidentally, this is not meant to imply that frameworks arising in one way or another from convex polyhedra are the only interesting or important ones from either a mathematical or a structural point of view—these are simply the only frameworks about which much is known.

For a framework $G(p)$ obtained from a convex polyhedron C by first adding new vertices in the interior of edges of C (so each edge of C is now subdivided by edges of $G(p)$) and then adding new noncrossing diagonal edges across each face of C, Alexandrov **[1**, Chapter 10] shows that $G(p)$ is stress free, i.e., rank $df(p)$ equals the number of edges of G. Therefore, for such frameworks, $G(p)$ is rigid (or, equivalently, infinitesimally rigid) in $\mathbb{R}^3$ if and only if $G(p)$ forms a triangulation of the surface of C. A version of Alexandrov's proof appears in Asimow and Roth **[3]**.

If new vertices are also allowed in the interior of faces of C before the addition of the new noncrossing edges in the faces of C, then the resulting framework may admit nontrivial stresses. A stress $\{\omega_{\{i,j\}}: \{i,j\}$ an edge of $G\}$ of a framework $G(p)$ is called a *facial stress* if there exists a face of C such that $\omega_{\{i,j\}}=0$ for all edges $[p_i,p_j]$ which do not lie in the face. Whiteley **[24]** proves a significant generalization of Alexandrov's result which states that, for such frameworks, every stress is a sum of facial stresses. Related results based on the study of infinitesimal flexings rather than stresses appear in Connelly **[13]**.

Finally, what happens if new vertices are present in the interior of faces of C before the faces of C are triangulated by noncrossing edges? In this case, the framework $G(p)$ is not infinitesimally rigid in $\mathbb{R}^3$ (to the vertices in the interior of a face, assign vectors which are perpendicular to the face and assign zero vectors to the remaining vertices in order to obtain an infinitesimal flexing), p is not a regular point of the edge function f, and $G(p)$ admits nontrivial stresses. Nevertheless, Connelly **[13]** shows that $G(p)$ is rigid in $\mathbb{R}^3$ by examining the second derivative of the preservation of edge length conditions. In other words, every triangulation of the surface of a convex polyhedron in $\mathbb{R}^3$ gives a rigid framework in $\mathbb{R}^3$.

References

1. A. D. Alexandrov, Konvexe Polyeder, Akademie–Verlag, Berlin, 1958.
2. L. Asimow and B. Roth, The rigidity of graphs, Trans. Amer. Math. Soc., 245 (1978) 279–289.
3. ______, The rigidity of graphs II, J. Math. Anal. Appl., 68 (1979) 171–190.
4. L. Auslander and R. E. MacKenzie, Introduction to Differentiable Manifolds, Dover, New York, 1977.
5. G. T. Bennett, Deformable octahedra, Proc. London. Math. Soc. (2), 10 (1912) 309–343.
6. E. D. Bolker, Bracing rectangular frameworks II, SIAM J. Appl. Math., 36 (1979) 491–508.
7. E. D. Bolker and H. Crapo, Bracing rectangular frameworks I, SIAM J. Appl. Math., 36 (1979) 473–490.
8. R. Bricard, Mémoire sur la théorie de l'octaèdre articulé, J. Math. Pures Appl. (5), 3 (1897) 113–148.
9. A. L. Cauchy, Sur les polygones et polyèdres, Second Mémoire, J. École Polytechnique, 19 (1813) 87–98.
10. R. Connelly, A counterexample to the rigidity conjecture for polyhedra, Inst. Hautes Études Sci. Publ. Math., 47 (1978) 333–338.
11. ______, A flexible sphere, Math. Intelligencer, 1 (1978) 130–131.
12. ______, The rigidity of polyhedral surfaces, Math. Mag., 52 (1979) 275–283.
13. ______, The rigidity of certain cabled frameworks and the second order rigidity of arbitrarily triangulated convex surfaces, Adv. in Math. (to appear).
14. M. Dehn, Über der Starrheit konvexer Polyeder, Math. Ann., 77 (1916) 466–473.

15. H. Gluck, Almost all simply connected closed surfaces are rigid, in Geometric Topology, Lecture Notes in Math., no. 438, Springer-Verlag, Berlin, 1975, pp. 225-239.

16. B. Grünbaum and G. C. Shephard, Lectures in lost mathematics, mimeographed notes, Univ. of Washington, Seattle.

17. D. Hilbert and S. Cohn-Vossen, Geometry and the Imagination, Chelsea, New York, 1952.

18. R. Kanayama, Bibliography on the theory of linkages, Tôhoku Math. J., 37 (1933) 294-319.

19. A. B. Kempe, On a general method of describing plane curves of the nth degree by linkwork, Proc. London Math. Soc., 7 (1876) 213–216.

20. N. H. Kuiper, Sphères polyédriques flexibles dans E^3, d'après Robert Connelly, Séminaire Bourbaki, 514 (Feb. 1978) 1–22.

21. G. Laman, On graphs and rigidity of plane skeletal structures, J. Engrg. Math., 4 (1970) 331–340.

22. J. Milnor, Singular Points of Complex Hypersurfaces, Ann. of Math. Studies, no. 61, Princeton University Press, Princeton, N. J., 1968.

23. H. Weyl, Über die Starrheit der Eiflächen und konvexen Polyeder, S.-B. Preuss. Akad. Wiss. (1917) 250-266.

24. W. Whiteley, Infinitesimally rigid polyhedra, preprint.

PLÜCKER EQUATIONS FOR CURVES

J. R. QUINE

Department of Mathematics, The Florida State University, Tallahassee, FL 32306

1. Introduction. For a smooth closed curve in the plane or the projective plane, the most evident topological characteristics are cusps and self-intersection points. There are several equations describing the number and nature of these. We will call these Plücker equations, after the ones given in the last century by Plücker for algebraic curves in complex projective 2-space. Although Plücker's original equations are part of algebraic geometry, we believe that, by looking at the analogous situation for smooth closed curves in real projective space, one can understand that the content of Plücker's equations is mainly topological. The situation in real projective space is easy to picture and makes a good starting point for understanding algebraic curves. We will begin this paper with the simplest Plücker theorem for curves in the plane, called the Umlaufsatz, and end with the classical Plücker equations.

One feature that we would like to emphasize is use of the dual correspondence. Plücker equations can be applied to the dual curve giving equations involving the number of inflection points and double tangents of the original curve. Our first goal will be to write a Plücker equation for real curves which can be applied to the dual curve and which will be a natural generalization of the Umlaufsatz. This will involve understanding the dual relationship between "winding number" and "tangent turning number." Once such a theorem is proved, we show that the classical Plücker equations can be proved in exactly the same way, with the help of some global analysis. We hope that the reader unfamiliar with techniques of global analysis can understand how they are a natural extension of the concept of the degree of a map.

2. Preliminaries. The real projective plane RP^2 can be thought of as the set of lines through

---

J. R. Quine received his Ph.D. at the University of Michigan in 1971 under the direction of F. W. Gehring. He had spent the preceding academic year teaching at Tuskegee Institute. Since 1971, he has been on the mathematics faculty of the Florida State University, where he is currently an associate professor; this article is based on a series of talks given there in the advanced topology seminar. He is interested in exploring connections between complex analysis, algebraic geometry, and differential topology.—*Editors*

the origin in $\mathbb{R}^3$. A line in RP^2 can be thought of as the set of lines through the origin in $\mathbb{R}^3$ contained in a plane. An alternative way of thinking of RP^2 is to see it as the unit sphere S^2 in $\mathbb{R}^3$ with antipodal points identified. A line in RP^2 is then a great circle. Now a nonzero point $x=(x_1,x_2,x_3)$ in $\mathbb{R}^3$ determines a unique line through the origin and therefore a point in RP^2. We say that x is a set of homogeneous coordinates for the point, often writing just "x" instead of "the point with homogeneous coordinates x." This brings up still another way to think of RP^2. Think of $\mathbb{R}^2$ as a subset of RP^2 by identifying the point (x_1,x_2) in $\mathbb{R}^2$ with the point $(x_1,x_2,1)$ in RP^2. The points in RP^2 not in $\mathbb{R}^2$ are then of the form $(x_1,x_2,0)$. These points are on a line called the line at infinity and denoted by l_∞. Thus we can think of RP^2 as $\mathbb{R}^2$ together with a line at infinity.

In projective geometry, we often use the dual correspondence, explained as follows: a line RP^2 determines a plane in $\mathbb{R}^3$; the normal to this plane determines a point in another copy of projective space denoted by RP^{2*}. This gives a one-to-one correspondence between lines in RP^2 and points in RP^{2*} called the dual correspondence. This correspondence can be described in another way. A line in RP^2, represented by $a_1x_1+a_2x_2+a_3x_3=0$ in $\mathbb{R}^3$, corresponds to the point $a=(a_1,a_2,a_3)$ in RP^{2*}. We say that $a\cdot x=0$ (dot product) is the equation of the line in RP^2 corresponding to the point a in RP^{2*}. Sometimes a is called a set of coordinates for the line $a\cdot x=0$. For example, $e_3=(0,0,1)$ in RP^{2*} is a set of coordinates for l_∞ in RP^2.

It is sometimes convenient to think of directed lines in RP^2. Indicating a direction on the line is the same as specifying a counterclockwise direction in the corresponding plane in $\mathbb{R}^3$ or giving the directed normal to that oriented plane. (This is the standard right-hand rule convention.) We say that a in $\mathbb{R}^3$ determines the directed line if it is in the direction of this normal. Thus we think of a and $-a$ being the same point in RP^{2*}, but determining different directions on the line $a\cdot x=0$. When we wish a to indicate a directed line, we will call it an oriented point in RP^{2*}.

There is a dual correspondence also between points in RP^2 and lines in RP^{2*}. A line in RP^{2*} is often thought of as a pencil of lines in RP^2. The set of lines $a\cdot x=0$ in RP^2 such that $a\cdot b=0$ is just the set (pencil) of lines through the point b. Likewise we can have an oriented pencil of lines, this being the same as an oriented point in RP^2.

The fundamental operations in projective space are finding the line joining two distinct points and finding the point of intersection of two distinct lines. By the dual correspondence described above, these are really the same operation. If a and b are distinct points in RP^2, the dual point to the line through a and b is denoted by $a\wedge b$. If a and b are thought of as vectors in $\mathbb{R}^3$, this is exactly the cross product. Likewise, if a and b are distinct points in RP^{2*}, $a\wedge b$ denotes the point of intersection of the lines $a\cdot x=0$ and $b\cdot x=0$. Again, this is just the cross product of two vectors.

We remark that all the definitions above are invariant under an orientation-preserving linear change of coordinates in $\mathbb{R}^3$.

3. Closed Curves in RP^2. Let S^1 be the unit circle in $\mathbb{R}^2$. A smooth map $f: S^1\to RP^2$ is called a closed curve. For picturing a closed curve, think of it as a curve in $\mathbb{R}^3$ which may cross l_∞ at a number of places. (See Fig. 1. The dotted lines are tangent lines to the curve at points on l_∞.) For viewing the curve crossing l_∞ we can change coordinates (project) so that l_∞ goes to a finite line.

We are interested in the following: (a) cusps of f, i.e., points $f(p)$ such that the differential of f at p is zero but the second-order derivatives are not all zero; (b) double points, i.e., points $f(p)$ such that $f(p)=f(q)$ with $p\neq q$ and $f(p)$ has exactly two preimages. In general we assume that the only singularities of f are cusps and the only multiple points are double points. Also we assume that the cusps are distinct from the double points and that the tangent lines at the double points are distinct. Every smooth curve is "close" to one satisfying these conditions.

Now we discuss the concept of a dual curve. Given a closed curve f, define a closed curve $\hat{f}$ by $\hat{f}(p)=b$ where $b\cdot x=0$ is the tangent line at $f(p)$. By tangent line at $f(p)$ we mean the limiting

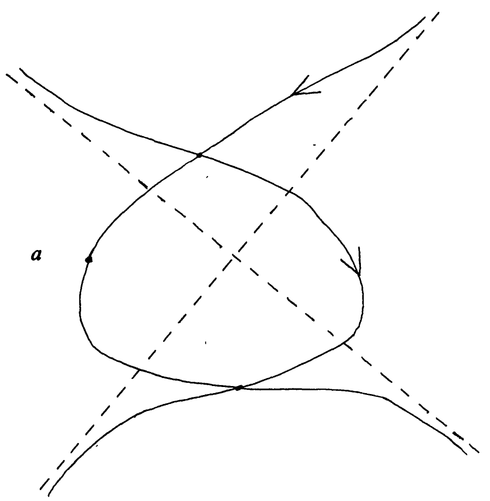


FIG. 1

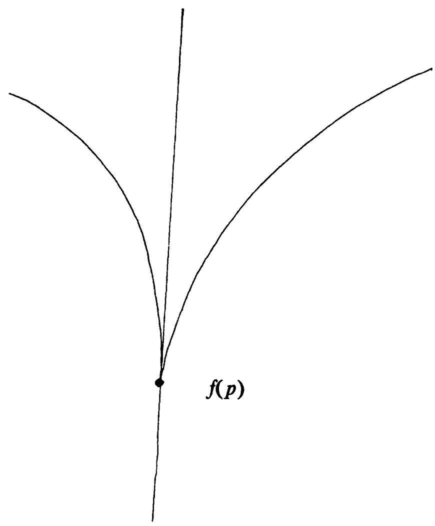


FIG. 2

position of the secant line $(f(p)\wedge f(q))\cdot x=0$ as q approaches p. Note that even at a cusp a tangent line is defined. (See Fig. 2.) The curve $\hat{f}$ is called the dual curve to f. If the dual curve $\hat{f}$ has cusps and double points, this can be seen in the behavior of the curve f. The point $\hat{f}(p)$ is a cusp of $\hat{f}$ if and only if $f(p)$ is an inflection point of f. In fact, this may be taken as a definition. If $\hat{f}(p)=\hat{f}(q)$ is a double point of $\hat{f}$ then the line $\hat{f}(p)\cdot x=0$ is called a double tangent of f. This line is tangent to f at exactly two points.

One can easily see that $\hat{\hat{f}}=f$ by observing that the secant line to $\hat{f}(p)$ and $\hat{f}(q)$ corresponds under the dual correspondence to the point $\hat{f}(p)\wedge\hat{f}(q)$. As p approaches q, $\hat{f}(p)\wedge\hat{f}(q)$ approaches $f(p)$.

Closely related to the dual curve is the secant map. All of the Plücker equations proved here will be proved by looking at the secant map. Let D be the set of pairs (p,q) such that $f(p)=f(q)$ and $p\neq q$. Let Δ be the diagonal, i.e., the set of pairs (p,q) such that $p=q$. The secant map s from $S^1\times S^1-(\Delta\cup D)$ to RP^{2*} is defined by $s(p,q)=f(p)\wedge f(q)$. Using the map $t\to(\cos 2\pi t,\sin 2\pi t)$ from $\mathbb{R}$ to S^1, we may think of the closed curve f as a map $f:I\to\mathrm{RP}^2$, where I is the unit interval $[0,1]$. Now think of s as a map from $I\times I-(\Delta\cup D)$ to RP^{2*}, where Δ and D are defined as above for p and q in I. It is more convenient to look at s restricted to $T-(\Delta\cup D)$, where $T\subset I\times I$ is the triangle formed by the set of pairs (p,q) such that $p\leqslant q$. If we let K be the set of points (p,p) on Δ such that $f(p)$ is a cusp, then s can be extended continuously to $T-(K\cup D)$ by defining $s(p,p)=\hat{f}(p)$ for (p,p) not in K. Thus the secant map restricted to Δ is essentially the dual curve.

4. Winding Numbers. If a is an oriented point in RP^2 and f is a closed curve, consider the map $f\wedge a$ from S^1 into the oriented line $a\cdot x=0$ in RP^{2*}. This oriented line is topologically the same as an oriented copy of S^1. Thus we may speak about the degree of the map $f\wedge a$ from S^1 to S^1. We denote this degree by $n(a)$ or $n(f,a)$. We remark that $f\wedge a$ is defined for all a, even if a is a point on the curve. At points p where $f(p)=a$, we define $f\wedge a$ to be the tangent line. We also remark that $n(f,a)$ is independent of the parametrization of f as long as the curve is traversed in the same direction.

If f is a curve in $\mathbb{R}^2$, then $n(f,a)$ has a familiar interpretation. If a is not on the curve, $n(f,a)$ is twice the winding number of f about a. If a is on the curve, $n(f,a)$ is twice the average of the

winding number about adjacent components of the complement of the curve. Likewise if b is an oriented point in RP^{2*}, and f is a curve in $\mathbb{R}^2$ without cusps, $n(\hat{f},e_3)$ has an interesting interpretation. Consider the Gauss map $g:S^1\to S^1$ sending p to the unit tangent of f at p. The degree of the Gauss map is sometimes called the tangent turning number, or rotation index. It happens that $n(f,e_3)$ is twice the tangent turning number. To see this, note that $\hat{f}(p)\wedge e_3$ is the intersection of l_∞ and the line through the origin parallel to the tangent line at p. Thus "winding number" and "tangent turning number" are the same because of the dual correspondence. It would be more correct, in fact, to speak of the tangent turning about an oriented line. The usual tangent turning number is then just the tangent turning number about l_∞ oriented by e_3.

5. The Umlaufsatz. The easiest Plücker equation to discuss is the Umlaufsatz (Hopf [7], Spivak [14], Chern [1], [2]). This theorem goes back to Riemann, but its modern proof is due to Hopf [7].

THEOREM 1 (Umlaufsatz). *If f is a closed curve in $\mathbb{R}^2$ without cusps or double points, then $n(f,e_3)=\pm 2$, i.e., the tangent turning number is ± 1.*

Proof. By reparametrizing and changing coordinates, we may assume that $f(0)=e_3$ and that the image of f in $\mathbb{R}^2$ lies in one of the half-planes determined by the tangent line to f at e_3.

We look at the map $s\wedge e_3$ sending the pair (p,q) to the intersection of the secant with the line at infinity. For $0\leqslant t\leqslant 1$, let γ_1 be the curve given by $(0,t)$, γ_2 be the curve given by $(1,t)$, and Δ be the curve given by (t,t). We may write $\partial T=\Delta-\gamma_1-\gamma_2$, where T denotes the boundary of the triangle T traversed in the counterclockwise direction. Since D and K are empty, s extends to all of T. Now $s\wedge e_3|_{\partial T}$, being a map from S^1 to l_∞ $(=S^1)$ has a well-defined degree. Since s extends to T, this degree must be zero. Now

$$\deg s\wedge e_3|_{\gamma_1}=\deg s\wedge e_3|_{\gamma_2}=\deg f\wedge e_3=n(e_3)$$

$$\deg s\wedge e_3|_{\Delta}=\deg\hat{f}\wedge e_3=\hat{n}(e_3)$$

where $\hat{n}(e_3)=n(\hat{f},e_3)$. Therefore

$$\deg s\wedge e_3|_{\partial T}=-2n(e_3)+\hat{n}(e_3)=0.$$

Since f lies in one of the half-planes determined by the tangent line to $f(0)$, we see that $n(e_3)=\pm 1$. The theorem now follows.

The Umlaufsatz can be generalized to allow for the presence of double points and cusps. Suppose $(p,q)\in D$. Let γ be the boundary traversed counterclockwise of a small square centered at (p,q). We can show that $\deg s\wedge e_3|_\gamma=\pm 2$. To see this, trace the progress of $s\wedge e_3$ in l_∞ as (p,q) traverses the boundary of the square. (See Fig. 3.) In fact, if v_1 and v_2 are the tangent vectors to the curve at p and q, respectively, then $\deg s\wedge e_3|_\gamma=-2\epsilon$, where $\epsilon=\pm 1$ gives the orientation of the pair (v_1,v_2). This may also be stated as follows: Let $\hat{f}(p)$ and $\hat{f}(q)$ give the tangent lines at p and q, oriented in the direction of motion; then

$$\deg s\wedge e_3|_\gamma=-\operatorname{sgn}\big(\hat{f}(p)\wedge\hat{f}(q)\big)\cdot e_3.$$

The latter has the advantage of using only our elementary operations in projective space.

Now write $\delta_{p,q}=\deg s\wedge e_3|_\gamma$ where $(p,q)\in D$ and γ is as above. Write $\delta=\Sigma\delta_{p,q}$, where the sum is over all $(p,q)\in D$. We note that δ depends on the parametrization, i.e., it depends on the choice of initial point $f(0)$. This is because $\operatorname{sgn}(\hat{f}(p)\wedge\hat{f}(q))\cdot e_3$ is reversed if p and q are reversed, and p and q are labeled by their relationship to the initial point. With this preparation we now have the following generalization of the Umlaufsatz.

THEOREM 2. *Suppose f is a closed curve in $\mathbb{R}^2$ without cusps. Suppose the initial point a is not a double point. Then $\delta=\hat{n}(e_3)-2n(a)$ where the orientation of a is chosen so that $e_3\cdot a>0$.*

Proof. Let B be a union of disjoint squares in T, each containing one point of D in its

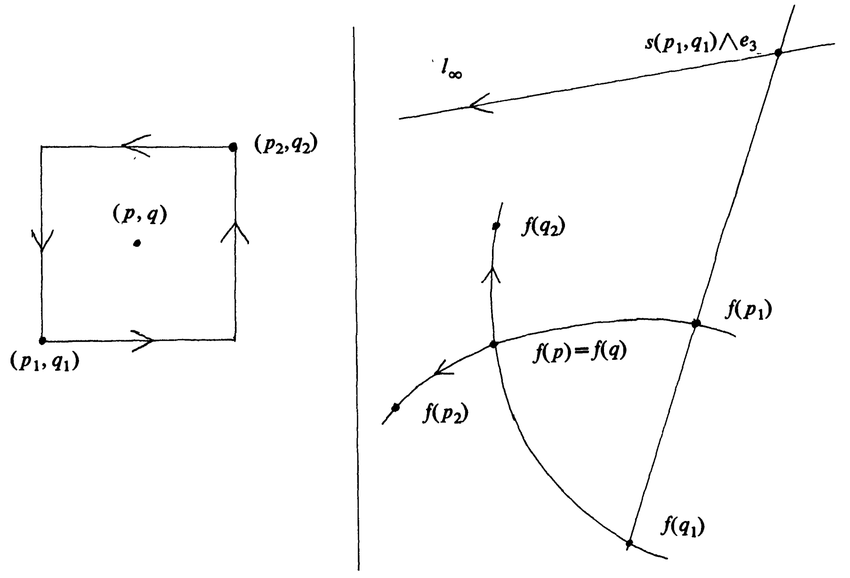


FIG. 3

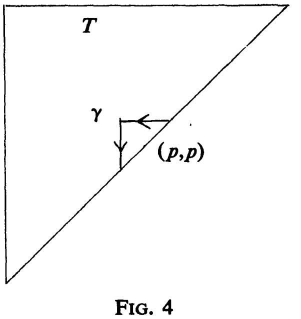


FIG. 4

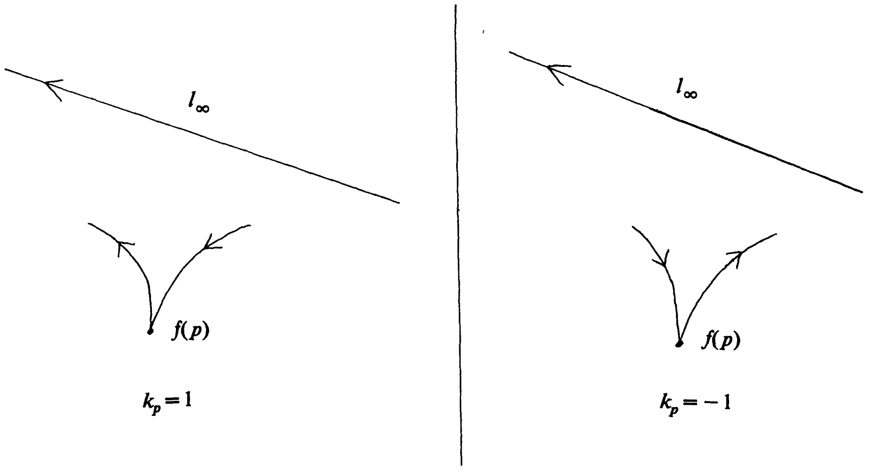


FIG. 5

interior. As before, look at the map $s\wedge e_3$ on T. By a simple homology argument

$$\delta = \deg s\wedge e_3|_{\partial B} = \deg s\wedge e_3|_{\partial T}.$$

Now with notation as before,

$$\deg s\wedge e_3|_{\gamma_1} = \deg(a\wedge f)\wedge e_3 = \deg f\wedge a,$$

where the latter equality depends on the fact that $e_3 \cdot a > 0$. Likewise deg $s\wedge e_3|_{\gamma_2} = \deg f\wedge a$. Since deg $s\wedge e_3|_\Delta = \hat{n}(e_3)$, the theorem follows.

Next we will generalize to allow for cusps. Suppose $f(p)$ is a cusp. Let γ_ε be the intersection of T with the boundary, oriented counterclockwise, of a small square having center (p,p) and sides ε. (See Fig. 4.) Although $s\wedge e_3|_{\gamma_\varepsilon}$ is not a closed curve in S^1, it does become closed as $\varepsilon\to 0$. We define the degree of this limiting curve to be k_p. The integer k_p will be ± 1, depending on the cusp. (See Fig. 5.) We note that k_p does not depend on the initial point. Now define $k = \Sigma k_p$, where the sum is over all p such that $f(p)$ is a cusp. A straightforward modification of the previous argument gives:

THEOREM 3. *Suppose f is a closed curve in $\mathbb{R}^2$. Suppose the initial point a is not a double point or cusp. Then $\delta + k = \hat{n}(e_3) - 2n(a)$, where the orientation of a is chosen so that $a\cdot e_3 > 0$.*

Theorems of this sort in the context of Euclidean space are proved in Hopf [7], Titus [15], and Whitney [17].

6. Generalization to the Projective Plane. Up to now all the theorems mentioned have been for curves in $\mathbb{R}^2$. The machinery of projective space would not be needed to prove these theorems. Our purpose, however, is to prove a generalization of the Umlaufsatz for curves in RP^2 which will be more closely related to the classical Plücker theorems. We will now modify the above proofs to allow for the curve crossing l_∞.

To modify the proof, we must look at pairs of points on l_∞. We will assume that l_∞ is not tangent to f at any point, i.e., e_3 is not on the curve $\hat{f}$. We will also assume that there are no cusps or double points on l_∞. Let S be the set of pairs (p, q) such that $f(p)\cdot e_3 = 0 = f(q)\cdot e_3$ and $p \neq q$. Note that if $(p, q)\in S$, then $s(p, q) = e_3$. Thus $s(p, q)\wedge e_3$ is undefined for $(p, q)\in S$. Now $s\wedge e_3$ defines a map from $T-(S\cup D\cup K)$ to l_∞ $(=S^1)$. Now we treat the points in S as we did the ones in D. Let γ be the boundary traversed counterclockwise of a small square with center $(p, q)\in S$. Define $\nu(p, q) = \deg\ s\wedge e_3|_\gamma$. This integer will be ± 2 depending on how f crosses l_∞ at p and q. By projecting properly, one of the two pictures in Fig. 6 will be correct. As

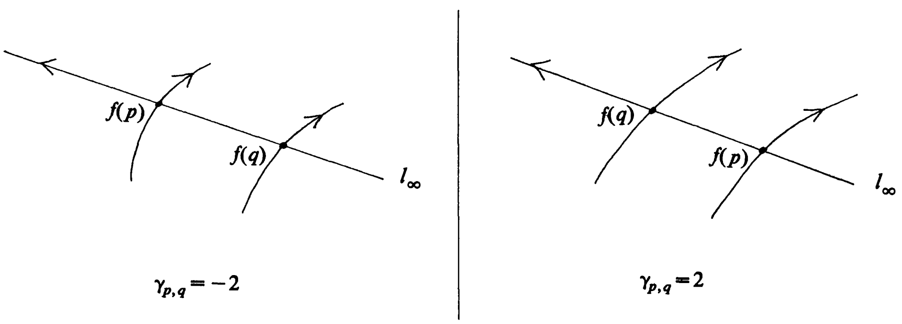


FIG. 6

is the case with points in D, we can show that deg $s\wedge e_3|_\gamma = 2\mathrm{sgn}(e_3\cdot \hat{f}(p)\wedge \hat{f}(q))$ where $\hat{f}(p)$ and $\hat{f}(q)$ give the oriented tangent lines. Now define $\gamma = \Sigma\gamma_{p,q}$ where the sum is over points p, q in S. If f crosses l_∞, say d times, then this is a sum of $d(d-1)/2$ terms. Now, by modifying the previous proof, we get:

THEOREM 4. *Suppose f is a closed curve in* RP^2. *Suppose the initial point a is neither a double point nor a cusp, and that no double point or cusp is on l_∞. Suppose that l_∞ is not tangent to the curve at any point. Then*

$$\nu+\delta+k=\hat{n}(e_3)-2n(a)$$

where the orientation of a is chosen so that $a\cdot e_3>0$.

In Fig. 1, for example, we have $\hat{n}(e_3)=0$, $n(a)=1$, $k=0$, $\delta=4$, and $\nu=-2$.

Note that Theorem 4 has the advantage over previous versions that it can be applied to the dual curve $\hat{f}$. Previous theorems deal only with curves in $\mathbb{R}^2$, whereas $\hat{f}$ will generally cross l_∞ ($e_3\cdot x=0$) in RP^{2*}. This happens when the tangent line passes through the origin in $\mathbb{R}^2$. As an example, the Umlaufsatz applied to $\hat{f}$ gives the following theorem: *A smooth curve in the projective plane with no double tangents or inflection points, and such that the tangent line never passes through the origin in $\mathbb{R}^2$, has $n(e_3)=\pm 2$.* For a more complete discussion, see Quine **[11]**. Our main purpose now is to discuss the relationship of Theorem 4 to Plücker's original equations.

7. Plücker's Original Equations. We now turn our attention to the original Plücker equations for algebraic curves in complex projective 2-space. Our purpose is to show that the methods of our previous proofs can be adapted to prove Plücker's equations. Our main tool again will be the topological degree of a map, this time from an oriented 2-manifold into the 2-sphere S^2. Also, as before, we use the secant map and the fundamental operations in projective space.

We think of complex projective space $\mathbb{C}\mathrm{P}^2$ as the set of complex lines in $\mathbb{C}^3$. We may think of it, as before, as $\mathbb{C}^2$ together with a line at infinity. An algebraic curve is given in $\mathbb{C}^2$ by an equation $F(x,y)=0$, where F is a polynomial in x and y with complex coefficients. More in line with our previous discussion, we will think of an algebraic curve as a map $f: M\to\mathbb{C}\mathrm{P}^2$, where M is a Riemann surface of genus g, and f is holomorphic. The map f is given on a neighborhood U in M by a holomorphic map $f_U: U\to\mathbb{C}^3-\{0\}$. The mapping f cannot be given globally by a holomorphic map from M into $\mathbb{C}^3-\{0\}$, and it is this fact that makes the algebraic case more interesting. It is true that f can be given by a nonvanishing map (section) $\phi: M\to L^3$ where L is a holomorphic line bundle over M. The reader unfamiliar with this terminology should think of ϕ as a map from M into a copy of $\mathbb{C}^3-\{0\}$ which varies from point to point.

The discussion of Plücker characteristics of an algebraic curve is similar to our previous one. Again, a cusp is just a singularity of the map f. We require that the first-order derivatives of f vanish, but not all second-order derivatives (in some coordinate system). A double point is where f crosses itself transversely. Here think of the image of a two-manifold crossing itself within a four-manifold. Another way to think is to think of f as a real algebraic curve and picture the real part of it. Then cusps and double points look the same as before.

The number $n(a)$ is defined as before as the degree of the map $f\wedge a$ from M into the line $a\cdot x=0$ in the dual space. This makes sense because lines in $\mathbb{C}\mathrm{P}^2$ are diffeomorphic to S^2. In complex projective space the orientation determined on S^2 by a and $-a$ are the same. Now $n(a)$ is the same for all points not on the curve, and this number d is called the order of the curve. The integer d can also be characterized as the degree of the polynomial F, or the number of times f crosses any line in $\mathbb{C}\mathrm{P}^2$. Now f has a dual $\hat{f}$ as before, which gives the coordinates of the complex tangent line at each nonsingular point. The order of $\hat{f}$ is denoted by d^*. Suppose now that the only singularities of f are cusps and that the image of f crosses itself only at double points. Let k be the number of cusps and δ the number of double points. Plücker's equations are

$$d^*+k+2-2g=2d \tag{1}$$

$$(d-1)(d-2)=2g+2k+2\delta. \tag{2}$$

If $\hat{f}$ is equally as nice as f, and if i is the number of inflection points of f, and τ the number of double tangents, we can apply (1) and (2) to the dual, getting

$$d+i+2-2g=2d^* \tag{3}$$

$$(d^*-1)(d^*-2)=2g+2i+2\tau. \tag{4}$$

We wish to sketch a proof of (2) which is analogous to the proof given for curves in RP^2. Classical proofs of (2) do not follow this line of reasoning, but use Bézout's theorem instead. See Griffiths and Harris [5]. Again, we look at the secant map $s: M\times M-(D\cup\Delta)\to\mathbb{C}\mathrm{P}^{2*}$ where Δ is the diagonal in $M\times M$ and D is defined as before to be the set of pairs (p,q) such that $f(p)=f(q)$ and $p\neq q$. Let K be the set of pairs (p,p) such that $f(p)$ is a cusp. Then s can be extended to $M\times M-(D\cup K)$ by defining $s(p,p)$ to be the point determined by the tangent line at p. Now let us assume that f intersects the line at infinity in d distinct points. By a proper choice of coordinates, this will be the case. Now let S, as before, be the set of pairs (p,q) such that $p\neq q$ and $f(p)\wedge f(q)=e_3$, i.e., $f(p)$ and $f(q)$ are on the line at infinity. Now look at the map $s\wedge e_3$ from $M\times M-(D\cup K\cup S)$ to the line at infinity. Let U be a small coordinate neighborhood in $M\times M$. In U, s is given by a map $s_U: U\to\mathbb{C}^3$; hence $s\wedge e_3$ is given by a map $v=s_U\wedge e_3: U\to\mathbb{C}^3$. Since the third coordinate is zero, we may consider it as a map into $\mathbb{C}^2$. Suppose now that U is a neighborhood of a point $(p,q)\in D$, and suppose that the local coordinates are given by (z,ζ) with $(0,0)$ corresponding to the point (p,q). Let B be the ball $|z|^2+|\zeta|^2\leqslant\epsilon$. Now v vanishes at $(0,0)$ and $v/|v|$ ($|v|=|v_1|^2+|v_2|^2$) defines a map from ∂B to the unit sphere $S^3\subset\mathbb{C}^2$. The degree of this map can be computed to be 1. Likewise for (p,q) in S a similar construction gives a map of degree 1, and for (p,q) in K the degree is 2. Now there are 2δ points in D and there are $d(d-1)/2$ pairs of intersection points with the line at infinity. Thus, combining all of these maps from boundaries of spheres containing points in $D\cup S\cup K$, we get a map of total degree $2\delta+2k+d(d-1)$.

It was at this point in our discussion for curves in RP^2 that we set the degree of the map analogous to the above map equal to the degree of the map on ∂T. Since $M\times M$ does not have a boundary, we must resort to some global analysis. In complex global analysis, the concept of degree of a map restricted to the boundary is replaced by the concept of Chern class. The map $s\wedge e_3$ is given locally by a holomorphic map from a neighborhood of $M\times M$ to $\mathbb{C}^2$, but globally it must be given by a section of L^2, where L is a line bundle over $M\times M$. The integer $2\delta+2k+d(d-1)$ counts the number of times this section vanishes. This must be equal to the integral over $M\times M$ of $c_2(L^2)$, the second Chern form of L^2. Now L can be expressed as a tensor product of familiar line bundles on $M\times M$ and, using standard computational methods for Chern forms, the integral of $c_2(L^2)$ can be computed to be $2((d-1)^2-g)$. Equating this to the expression above, we get (2).

We mention equation (1) only briefly. It can be proved by a similar analysis of L restricted to Δ in $M\times M$. In this case, we compute the integral over Δ of $c_1(L)$ to be $d+2g-2$, and the number of zeros of a section of L to be d^*+k. The basic idea is that all of the Plücker equations can be obtained by essentially equating the degree of two homologous maps.

In closing, we mention that there are noncompact versions of (1) and (2). By this, we mean that we consider a noncompact Riemann surface M and holomorphic map f and see how the above reasoning generalizes. A noncompact version of (1) is given in Weyl and Weyl [16] and is nicely explained in Cowen and Griffiths [3]. It relates to Nevanlinna theory of holomorphic curves. A non-compact version of (2) is given in Quine and Yang [12]. Some other theorems dealing with Plücker characteristics of real curves are the theorem of Halpern [6] and Fabricius-Bjerre [4], and the three-inflection-point theorem [9], [13]. The techniques of proof in these theorems of Fabricius-Bjerre and Halpern, although similar to the above, are enough different that they could not be dealt with quickly here. Also, it is an interesting question whether the three-inflection-point theorem fits with these other theorems at all, even though it deals only with Plücker characteristics. It does not appear to follow from any set of Plücker equations for smooth curves. (See Pohl [9].) It would be nice to see a unified theory which included all of these theorems.

One final remark: This paper began with the Umlaufsatz and proceeded through several generalizations to Plücker's equations for algebraic curves. What is the Umlaufsatz for algebraic curves? The natural generalization of a simple closed curve is an algebraic curve without singularities, i.e., with $k=0$ and $\delta=0$. For such a curve, (1) and (2) give $d^*=d(d-1)$. This familiar fact from algebraic geometry may thus be thought of as a direct descendant (logically) of the Umlaufsatz.

References

1. S. S. Chern, Curves and Surfaces in Euclidean Space, MAA Studies in Mathematics, vol. 4, 1967.

2. ______, From triangles to manifolds, this MONTHLY, 86 (1979) 339–349.

3. M. Cowen and P. A. Griffiths, Holomorphic curves and metrics of negative curvature, J. d'Analyse Math., 29 (1976) 93–153.

4. Fr. Fabricius-Bjerre, A relation between the number of singular points and singular lines of a plane closed curve, Math. Scand., 40 (1977) 20–24.

5. P. A. Griffiths and J. Harris, Principles of Algebraic Geometry, Wiley, New York, 1978.

6. B. Halpern, Global theorems for closed plane curves, Bull. Amer. Math. Soc., 76 (1970) 96–100.

7. H. Hopf, Über die Drehung der Tangenten und Sehnen ebener Kurven, Compositio Math., 2 (1935) 50–62.

8. L. H. Kauffman and T. F. Banchoff, Immersions and mod-2 quadratic forms, this MONTHLY, 84 (1977) 168–185.

9. W. F. Pohl, Singularities in the Differential Geometry of Submanifolds, Lecture Notes in Mathematics, no. 209, Springer-Verlag, New York, 1971.

10. J. R. Quine, Plücker relations for $p(e^{i\theta})$, Math. Scand., 44 (1979) 289–294.

11. ______, A Plücker-type equation for curves in real projective space (to appear).

12. J. R. Quine and P. Yang, A non-compact version of Plücker's second equation, Duke J. Math., 47 (1) (1980).

13. S. Sasaki, The minimum number of points of inflection of closed curves in the projective plane, Tôhoku Math. J., 9 (1951) 113–117.

14. M. Spivak, Differential Geometry, vol. 2, Publish or Perish, Boston, 1970.

15. C. J. Titus, A theory of normal curves and some applications, Pacific J. Math., 10 (1960) 1083–1096.

16. H. Weyl and J. Weyl, Meromorphic Functions and Analytic Curves, Princeton University Press, Princeton, N.J., 1943.

17. H. Whitney, On regular closed curves in the plane, Compositio Math., 4 (1937) 276–284.

ARITHMETIC WITH ROMAN NUMERALS

JAMES G. KENNEDY
10479 Holman Avenue, Los Angeles, CA 90024

It is widely believed that calculation with Roman numerals is a primitive process, so clumsy that operations beyond simple addition and subtraction become almost impossible. A recent article [1] in this MONTHLY described it as a degenerate system representing only the natural numbers, and the characters V, X, etc., as being only shorthand notation for strings of ones. On the contrary, the Roman system incorporates features permitting algorithms that make multiplication and division even easier than in our place value decimal system, as will be demonstrated here.

Roman numerals apparently did develop from a primitive tallying system using strings of ones up to nine, and other symbols for powers of ten. The Etruscans are credited with introducing the symbols V, L, and D for 5, 50, and 500 to reduce the number of tallies [2]. In so

The author received a B.S. in Electrical Engineering in 1949 from Illinois Institute of Technology. Most of his career has been spent with Hughes Aircraft Company, working on advanced radar and electro-optical systems; he is currently involved with computer controlled automatic test equipment for these systems. He points out that his is the only nonteaching field in which professionals study arithmetic. This is because the design of efficient algorithms is of prime importance not only in computer systems themselves but also in programs that solve numerical problems. Consequently he is interested in all types of novel computational methods.—*Editors*

doing they created a practical compromise between the pure tallying system and the Arabic/Hindu system with its separate symbols for zero through nine. Since each Roman symbol uniquely represents a number, the zero of positional systems is unnecessary. Arithmetic was commonly done by Romans on the abacus or counting table. What methods they used for written calculations, if any, have apparently been lost, and the methods described here may be a modern invention.

Multiplication. To demonstrate the multiplication process we first set up a table with successive characters heading each column. To multiply CXXI by XVI we write the pattern of the multiplicand CXXI at the top for reference, using a tally mark T for each character in the appropriate column (see Example 1). The multiplier is written below it on the heavy horizontal line.

To find each partial product, shift the multiplicand product to the left two columns for X, one for V, and none for I. The process can be simply expressed as: shift the multiplicand pattern until the units position falls under the multiplier character. Finally, the T's are summed by column and replaced by Roman characters, after which the usual consolidating is done, giving an answer here of MCMXXXVI, or $121 \times 16 = 1936$.

M	D	C	L	X	V	I
		T		TT		T
				T	T	T
		T		TT		T
	T		TT		T	
T		TT		T		
M	D	CCC	LL	XXX	V	I

EXAMPLE 1

A separate shift rule is needed for multiplying any Etruscan character by another. Otherwise, the method above would yield X for V times V. In this case the same shift is used, but the tally is written twice, plus once more in the column to the right. Because of this it is more convenient to tally Etruscan numbers with lower case t's, although mathematically there is no significance to the difference. Example 2 demonstrates the use of both the basic shift and the Etruscan shift, multiplying XVI by VI to obtain LXXXXVI.

The system accommodates Roman subtractive notation equally well. Example 3 illustrates taking the product of XLIV by XLIV, or 44^2. The previous rules are used with the following

L	X	V	I	
	T	t	T	
		t	T	
T	tt	tT		
	T	t	T	
L	XXX	VVV	I	=LXXXXVI

EXAMPLE 2

M	D	C	L	X	V	I
			t	T′	t	T′
			t	T′	t	T′
			t′	T	t′	T
		tt	tT′	tt	tT′	
	t′	T	t′	T		
tt	tT′	tt	tT′			
MM	D′	CCCCC	L′L′	XXXX	V′	I

=MCMXXXVI.

EXAMPLE 3

adjustments for negative X's and I's:

(1) A negative is represented by a primed T and located in its normal column, e.g., a T′ in the X column represents a negative X;

(2) The partial product of an unprimed multiplier character is shifted and written without change;

(3) The partial product of a primed multiplier character is shifted and written with primes added to unprimed T's and with the primes removed from the primed T's;

(4) Tallies are added algebraically by column, and the sums are converted to Roman characters.

The counting table format, together with the basic shift, the Etruscan shift, and the rules for negatives completely define multiplication.

	C	L	X	V	I	
			T	t		
(a)			TT	t		=XXV (Quot.)
(b)	TT~~T~~	(TT)T	TT	T	TT	
(c)	TT	t t				
		T	TT	T		
(d)		T	tt	t		
(e)					TT	=II (Rem.)

Line (b) shows the Roman form of borrowing. The borrowed T is crossed out and the equivalent value appears in parentheses to the right.

Line (c) shows double subtraction.

Line (d) shows multiplication of Etruscan characters.

Line (e) shows the remainder of II.

EXAMPLE 4

Division. Long division can be performed by analogy using the following procedure. To divide CCCLXXVII by XV as shown in Example 4, write the divisor pattern for reference as before at the top, the dividend pattern in line (b), and the quotient in line (a). A partial divisor is formed by shifting the divisor pattern as far to the left as possible with respect to the dividend. The quotient character is written in the same column as the unit position of the shifted pattern, which determines whether it is Etruscan. The divisor is written beneath the dividend once if the quotient is Etruscan, or from one to four times if it is not (in line (c) it is used twice). The lines are then subtracted, and the process is repeated until the remainder is less than the quotient.

These examples demonstrate clearly the remarkable simplicity of multiplication and division with Roman numerals. Key features are:

—Although place value is not used in the numeral system, it is introduced in the counting table just as it is in the abacus. The combination of numerals with the table forms a place value system with a repeating multiplicative pattern of 5-2-5-2.

—The explicit zero is not needed in the numeral system or in the calculating table with its fixed columns.

—Most important, the algorithms *do not* require a knowledge of the multiplication tables. Sequential digit systems with place values require tables for calculating, but Roman numerals with their repeating pattern require only the two shift rules in the calculating table.

I wish to express my thanks to Jane I. Robertson for her generous help and suggestions during the preparation of this paper, and also to A. Cooperband for comments and help in developing the final forms of the algorithms.

References

1. Jane I. Robertson, How to do arithmetic, this MONTHLY, 86 (1979), 431–439.
2. L. Hogben, Mathematics for the Million, W. W. Norton, New York, 1967.

AN APPLICATION OF POINCARÉ'S RECURRENCE THEOREM TO ACADEMIC ADMINISTRATION

KENNETH R. MEYER
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221

The present trend in science is to apply classical mathematics to nontraditional areas. This note gives an application of a classical theorem of dynamical systems to a long neglected area of study, academic administration, and thus proves that scientific research and academic administration are not mutually disjoint. It is the author's hope that other administrations will apply their early training in scientific research to study the quagmire into which they have slipped and thus carry forth this work.

A recurrent orbit in a system is one that returns infinitely often arbitrarily close to its initial position. Poinçaré's recurrence theorem [1] states: *In a compact conservative system almost all orbits are recurrent.* Poincaré discovered this theorem in his investigations into the motion of celestial bodies, and until now it has not had applications to such terrestrial matters as academic administrative structures. However, we shall show that this theorem can easily be applied to explain an often observed phenomena.

LEMMA 1. *An academic administrative system is conservative.*

Proof. All decisions are made by applying the principle of least action and therefore the system is conservative by a classical theorem of Maupertuis [2].

LEMMA 2. *An academic administrative system is compact.*

Proof. The system is governed by a finite number of arbitrarily short-sighted deans and is compact by definition.

Lemmas 1 and 2 verify the hypothesis of Poincaré's recurrence theorem and therefore the conclusions hold for all academic administrations. An immediate consequence of this result is:

THEOREM 1. *Almost all administrators vacillate.*

Finally, since many conservative systems are reversible, an administrator will not only return infinitely often to the same position but must have been there infinitely often in the past.

This paper was presented at a meeting of the chairmen of the Ohio State University system in Columbus, Ohio, October 1978.

References

1. H. Poincaré, Les Méthodes Nouvelles de la Mécanique Céleste, Gauthier-Villars, Paris, 1892.
2. E. T. Whittaker, A Treatise on the Analytic Dynamics of Particles and Rigid Bodies, Cambridge University Press, 1904.

SUM-PRESERVING REARRANGEMENTS OF INFINITE SERIES

PAUL SCHAEFER
Department of Mathematics, SUNY, College of Arts and Science at Geneseo, Geneseo, NY 14454

1. Introduction. Every student of advanced calculus knows that an absolutely convergent series of real numbers may be rearranged in an arbitrary fashion to obtain a new series which converges to the same sum as that of the original series. The student also finds that conditionally convergent series behave somewhat differently in this regard. Indeed, Riemann proved that such series can be rearranged to converge to any arbitrary real number, or even to diverge. Moreover, it has been shown by J. H. Smith **[8]** that for any conditionally convergent real series and any real number, there is a rearrangement of a prescribed "cycle type" which converges to that number.

Yet, there obviously are rearrangements which *do* preserve the convergence and sum of all infinite series, whether they converge absolutely or merely conditionally. For example, if the series, $u_1+u_2+u_3+\cdots$, converges, then the rearrangement, $u_2+u_3+u_1+u_5+u_6+u_4+\cdots$, is easily seen to converge to the same sum. It would seem reasonable to try to characterize those rearrangements of series which preserve sums of convergent series. This paper surveys the several approaches to the problem to date and gives another characterization of such rearrangements. For the convenience of would-be series rearrangers, five somewhat simpler sufficient but not necessary conditions for "sum-preserving" rearrangements are also developed. The reader is invited to add to this list.

Rearrangements of series can be described in terms of permutations of the positive integers. Let N denote the set of all positive integers. A *permutation* p of N, of course, is a one-to-one mapping of N onto itself. Let p_j be the image of j under the permutation p. The series Σu_{p_j} is

Paul Schaefer received his Ph.D. from the University of Pittsburgh under George Laush in 1963. He has held faculty positions at the Rochester Institute of Technology and SUNY at Albany, and participated in a faculty exchange at California State University at Los Angeles. Since 1967 he has been Professor of Mathematics at SUNY College at Geneseo. His research interests and publications are in the area of series and summability.—*Editors*

then a *rearrangement* of the series Σu_j. Such rearrangement is said to be *sum-preserving* if the former series converges to s whenever the latter one does. In the next three sections, necessary and sufficient conditions for sum-preserving rearrangements are given. The reader might find it instructive to check each of them against the example given above.

In studying questions of rearrangements of series, two essentially different techniques are used. One, which might be called combinatorial, relies heavily upon properties of N and permutations. The other uses results from summability theory about infinite matrix transformations of sequences or series into sequences or series. Both approaches will be illustrated in this paper. Combinatorial methods, although considerably less elegant, are perhaps more suggestive (one can draw pictures!).

2. Levi's Condition. In 1946, F. W. Levi [6] developed the first criterion for a rearrangement to be sum-preserving. Let p be the permutation inducing the rearrangement of a series. For each positive integer n, let s_n be the nth partial sum of the series Σu_j, and let t_n be the nth partial sum of the rearranged series Σu_{p_j}. Observe that the term u_j appears in t_n if and only if $j \in \{p_1, \ldots, p_n\}$, or, equivalently, if and only if $p^{-1}(j) \leqslant n$. Levi calls a term u_j of the original series a "jumping out term" for n if and only if $1 \leqslant j \leqslant n$ and $p^{-1}(j) > n$. These terms appear in s_n but not in t_n. A term u_j is a "jumping in term" for n exactly when $j > n$ and $p^{-1}(j) \leqslant n$, so that such terms appear in t_n but not in s_n. Then, $t_n = s_n + x_n - y_n$, where x_n is the sum of all jumping in terms for n, and y_n is the sum of all jumping out terms for n. A block of consecutive terms appearing in Σu_j is called a "bunch" for n if the block contains only jumping in terms for n or jumping out terms for n, *and* the block is maximal with respect to this property. Let $B(n)$ be the number of bunches for n. With this terminology, Levi's criterion can be stated as follows.

THEOREM 1. *A rearrangement is sum-preserving if and only if for the permutation p inducing the rearrangement there is a positive integer M so that $B(n) \leqslant M$ for each positive integer n.*

Levi's proof shows that when p is any permutation with $B(n) \leqslant M$ for all n, if $s_n \to s$, then $|x_n - y_n| \to 0$, so $t_n \to s$ also. On the other hand, he shows that for any permutation p with $\{B(n)\}$ not uniformly bounded, there is a convergent series whose rearrangement by p does not converge to the sum of the original series.

The proof given by Levi uses combinatorial methods throughout. In 1966, U. C. Guha [4] gave a succinct proof of Levi's theorem using a result of Bosanquet about series-to-sequence (actually, series-to-function) transformations by infinite matrices.

3. Agnew's Condition. Using techniques of summability theory, R. P. Agnew [2] in 1955 found another necessary and sufficient condition for a rearrangement to be sum-preserving. (A similar condition was developed recently by P. A. B. Pleasants [7].)

THEOREM 2. *The rearrangement induced by the permutation p is sum-preserving if and only if there is a positive integer M so that for each positive integer j, the set $\{p_1, \ldots, p_j\}$ is representable as the union of M or fewer blocks of consecutive integers.*

Theorem 1 and Theorem 2 are equivalent, as the following discussion shows. For any permutation p of N and any positive integer n, let O_n be the set of indices j which correspond to jumping-out terms u_j for n, and let I_n be the set of j's corresponding to jumping-in terms u_j for n. The number of bunches for n, $B(n)$, is the number of blocks of consecutive integers whose union is $O_n \cup I_n$. Since $\{p_1, \ldots, p_n\} = ([1, n] - O_n) \cup I_n$, the equivalence of Levi's and Agnew's conditions is readily apparent.

Agnew's proof was based upon regular matrix transformations of sequences into sequences. Later in this paper, a combinatorial proof of Agnew's result is given. It will be seen that this latter proof carries over easily to series in Banach spaces.

It should be pointed out that Pleasants considers rearrangements of series somewhat differ-

ently than is done here. In his paper, the permutation p moves the term u_j to where the term u_{p_j} was in the original series. This has the effect of interchanging the roles of the permutation p and its inverse in some of his theorems relative to those given here.

4. Another Condition. One can obtain another characterization of sum-preserving rearrangements in the following way. Let p be the permutation of N which induces the rearrangement. For each $k \in N$, let m_k be the smaller of $p^{-1}(k)$ and $p^{-1}(k+1)$, and let M_k be the larger of these two numbers. Set $I_k = \{j | j \in N \text{ and } m_k \leqslant j < M_k\}$. Note that since p is one-to-one, $I_{k_1} \neq I_{k_2}$ whenever $k_1 \neq k_2$. The criterion can be given in the following way.

THEOREM 3. *The rearrangement induced by the permutation p is sum-preserving if and only if there is a positive integer K so that every collection of K intervals,* $\{I_{k_1}, I_{k_2}, \ldots, I_{k_K}\}$, *has an empty intersection.*

This result was suggested after consideration of sum-preserving series-to-series matrix transformations. It will be shown later in this paper by combinatorial methods to be equivalent to Theorem 2. A summability proof will also be given so as to illustrate this kind of approach to the problem.

5. Some Elementary Remarks About Permutations. In order to investigate rearrangements more closely, let us make a few observations about permutations of the positive integers.

Let p be a permutation of N and let p_k be the image of the positive integer k under the permutation p. Clearly, $p_k \to \infty$ as $k \to \infty$. Since p is a permutation, it is easy to see that, if n is any positive integer, there is a K so that $n \in \{p_1, \ldots, p_k\}$ for all $k \geqslant K$. Furthermore, for any n we have $\{1, \ldots, n\} \subset \{p_1, \ldots, p_k\}$ for all sufficiently large k.

Any finite subset of N can be represented as the union of a disjoint collection of blocks of consecutive integers. We shall use the notation $J = [c, d]$ for such intervals; so $[c, d] = \{x | x \in N$ and $c \leqslant x \leqslant d\}$, admitting the possibility that $c = d$ as well. If p is any permutation of N, then, corresponding to each $j \in N$, there is a disjoint collection, $\{J(j, 1), \ldots, J(j, n_j)\}$ of intervals such that the set $\{p_1, \ldots, p_j\}$ is given by $J(j,1) \cup J(j,2) \cup \cdots \cup J(j, n_j)$. Agnew's criterion in Theorem 2 is that for the permutation p of N the set of integers $\{n_j\}$ is bounded.

Let us arrange the notation so that the interval $J(j, m)$ is to the left of $J(j, n)$ whenever $m < n$. In all of the following discussion, it will be assumed that j is large enough to have $1 \in \{p_1, \ldots, p_j\}$. The interval $J(j, 1)$ will always be of the form $[1, b(j)]$ for some integer $b(j)$, where $1 \leqslant b(j) \leqslant j$. The right-hand endpoint of $J(j, n_j)$ will be denoted by $B(j)$, the largest element of the set $\{p_1, \ldots, p_j\}$. Thus, $B(j) \geqslant j$.

6. Sufficient Conditions for the Preservation of Sums. Before getting to the proofs of Theorems 2 and 3, let us give five simple conditions for a rearrangement to preserve sums.

The permutation given in the example mentioned in the Introduction has the property that it does not move any integer very far from its original location in N. In fact, $|p_j - j| \leqslant 2$ for all j. This suggests the following criterion.

CONDITION 1. *There is a positive integer B so that* $|p_j - j| \leqslant B$ *for all j.*

In fact, as the discussion below shows, a weaker one-sided condition is sufficient for a permutation to produce a sum-preserving rearrangement of a convergent series.

CONDITION 2. *There is a positive integer B so that* $p_j \leqslant j + B$ *for all j.*

Condition 2 implies that, for each j, the numbers $p_1, \ldots, p_j$ are located somewhere in the interval $[1, j+B]$. We estimate the maximum number of disjoint intervals into which $\{p_1, \ldots, p_j\}$ could be decomposed. If $j + B = 2q$, an even integer, then the worst possible locations of the p_i's (the ones yielding the most intervals) would be to have q of them at $1, 3, 5, \ldots, (j+B-1)$, or at $2, 4, \ldots, (j+B)$. One of the $(j-q)$ remaining p_i's might be located at either end of $[1, j+B]$, but

all of the other remaining ones must fall somewhere in the gaps in the first q locations. Thus, the maximum number of intervals in the representation of $\{p_1,\ldots,p_j\}$ when $j+B=2q$, cannot exceed $q-(j-q-1)=B+1$. Similar considerations for $j+B=2q-1$ also yield a bound $B+1$. This shows that, in the notation of the preceding section, $n_j\leqslant B+1$ for all j, so the hypothesis of Theorem 2 is satisfied.

EXAMPLE 1. The permutation p which takes positive integral powers of 2 into their halves and which maps the remaining positive integers onto themselves consecutively satisfies Condition 2 but not Condition 1. We have $p_j=2^{n-1}$ if $j=2^n$, $p_j=2^n+1$ if $j=2^n-1$ and $p_j=j+1$ otherwise. Since $|p_j-j|=2^{n-1}$ when $j=2^n$, Condition 1 fails to hold. But it is easy to see that $p_j\leqslant j+2$ for all j.

Both Levi and Pleasants have observed that the set of all permutations which induce sum-preserving rearrangements of series is not a group. See Pleasants [7] for an example of a sum-preserving rearrangement induced by a permutation whose inverse does not have this property. (Keep in mind that Pleasants's notation differs from that of the other authors mentioned here.) However, Condition 1 is sufficient to assure that both the permutation p and its inverse, p^{-1}, induce sum-preserving rearrangements, for then, both permutations would satisfy Condition 2.

Further discussion of the algebraic structure of the set of convergence-preserving rearrangements of series can be found in **[9]**. In addition to Condition 1 above, the author mentions two other conditions which are sufficient for both the permutations p and its inverse p^{-1} to be convergence-preserving: (A) $\lim(p_j)/j=1$, and (B) $0<\inf(p_j)/j$ and $\sup(p_j)/j<+\infty$. Clearly, Condition 1 implies (A), which in turn implies (B). See also **[10]**.

It might be suspected that some restriction on the spread of the elements in $\{p_1,\ldots,p_j\}$ would be useful. Such a restriction, in the notation of Section 5 above, might be the following one.

CONDITION 3. *There is a positive integer B so that $B(j)-b(j)\leqslant B$ for all j.*

This condition, in fact, implies Condition 2, since $p_j\leqslant B(j)$ and $b(j)\leqslant j$, but the example given above shows that it is not equivalent to it. For Example 1, when $j=2^n-1$ we have $B(j)-b(j)=(2^n+1)-(2^{n-1}-1)=2^{n-1}+2$.

Condition 3, while being weaker than Condition 2, suggests another approach to the problem at hand. It is possible to give a direct proof that Condition 3 implies that the permutation p yields a sum-preserving rearrangement by observing that for any j,

$$\sum_{k=1}^{j} u_{p_k}=\sum_{k=1}^{b(j)} u_k+\sum{}^{*} u_{p_k},$$

where the second summation on the right is over those p_k's which satisfy $b(j)+1<p_k\leqslant B(j)$. Since $b(j)\to\infty$ as $j\to\infty$, and since $u_k\to 0$ if $\Sigma\, u_k$ converges to s, corresponding to any $\varepsilon>0$ one can simultaneously make the first sum on the right differ from s by less than $\varepsilon/2$ in absolute value, and make each term in the second sum less than $\varepsilon/2B$ in absolute value for all $p_k>b(j)+1$ when j is sufficiently large, and complete the proof in the usual way. The key to the success of this approach is *not* the size or growth rate of the $B(j)$'s or p_j's, but rather the cardinality of the set of indices appearing in the second summation, Σ^*. This cardinality is, of course, $j-b(j)$.

CONDITION 4. *There is a positive integer C so that $j-b(j)\leqslant C$ for all j.*

Let us also show directly that Condition 4 implies that Agnew's criterion is satisfied. If the intervals $J(j,k)$ in the decomposition of $\{p_1,\ldots,p_j\}$ are written as $J(j,k)=[c_k,d_k]$, $k=2,\ldots,n_j$, their cardinality is (d_k-c_k+1). Since $\{p_1,\ldots,p_j\}$ is the union of (n_j-1) of these disjoint intervals and the interval $[1,b(j)]$, it follows that

$$j=b(j)+\sum_{k=2}^{n_j}(d_k-c_k+1)=b(j)+(n_j-1)+\sum_{k=2}^{n_j}(d_k-c_k)\geqslant b(j)+(n_j-1).$$

If Condition 4 holds, we have $n_j\leqslant C+1$ for all j.

EXAMPLE 2. The permutation which is the inverse of the one given in Example 1 satisfies Condition 4 but not Condition 2. Thus, $p_1=2$, $p_j=2^{n+1}$ when $j=2^n$, $p_j=2^n-1$ if $j=2^n+1$, and $p_j=j-1$ otherwise. Since $p_j=2^n+2^n$ when $j=2^n$, Condition 2 is not met, but it can be seen easily that $j-b(j)\leqslant 2$ for all $j\geqslant 3$.

Not surprisingly, Condition 4 is not necessary for the preservation of sums.

EXAMPLE 3. Let the permutation p be given by the following scheme:

$$N:1\ 2\ 3\ 4\quad 5\ 6\ 7\ 8\ 9\ 10\ 11\ldots$$

$$p(N):1\ 4\ 3\ 2\ 10\ 9\ 8\ 7\ 6\quad 5\ 22\ldots$$

Condition 4 does not hold, since $j-b(j)$ is unbounded. (Take $j=3,9,21,\ldots,3\cdot 2^q-3,\ldots$ to see this.) However, it is obvious that $\{p_1,\ldots,p_j\}$ is the union of at most *two* disjoint intervals for each j.

Let us formulate another sufficient condition, this time basing it on Theorem 3. A simple combinatorial argument shows that every sufficiently large positive integer j belongs to exactly $(m+1)(m+2)/2$ different intervals of positive integers whose lengths do not exceed a fixed positive integer m. Thus, for $j>m$, j is in the $(m+1)$ intervals $[j,j],[j,j+1],\ldots,[j,j+m]$, and in the m intervals $[j-1,j],[j-1,j+1],\ldots,[j-1,j-1+m]$, and in the $(m-1)$ intervals $[j-2,j]$, $\ldots,[j-2,j-2+m],\ldots$, and finally in the single interval $[j-m,j]$.

Now, suppose that for the permutation p, all of the intervals I_k defined by p as in Theorem 3 are such that their lengths are bounded by m, say. For any positive integer j, with $j>m$, the collection of all I_k's to which it belongs must be a subset of the set of intervals considered in the preceding paragraph. Hence, any collection of $K=1+(m+1)(m+2)/2$ such intervals must have an empty intersection. These considerations lead to the next condition.

CONDITION 5. *There is a positive integer m so that the lengths of all intervals I_k are bounded by m.*

This, however, is merely a sufficient condition for preservation of sums. For the permutation of Example 3, one can see that the lengths of I_k when $k=4,10,22,46,\ldots$ increase without bound. In fact, the length of I_k when $k=3\cdot 2^q-2$ is $9\cdot 2^{q-1}-2$ for $q\geqslant 1$.

7. The Equivalence of Theorem 2 and Theorem 3. Let us now investigate the relationship between the conditions of Theorem 2 and Theorem 3, in order to show that they are equivalent.

A permutation which is different from the identity mapping obviously cannot preserve the order of N. There are, however, some useful relationships between the ordering of elements of N and the images of elements of N under a permutation p. For example, if q and r are positive integers, then the assertion that $q\leqslant r$ is equivalent to saying that $p_q\in\{p_1,\ldots,p_r\}$, and this is equivalent to the statement $p_r\in\{p_q,p_{q+1},\ldots\}$. Similar assertions hold for strict inequality.

Let p be a permutation of N. Recall that for Theorem 3, $I_k=[m_k,M_k-1]$, where m_k is the smaller of $p^{-1}(k)$ and $p^{-1}(k+1)$, and M_k is the larger of these two numbers. Note that a positive integer j belongs to I_k if and only if either (1) $k\in\{p_1,\ldots,p_j\}$ and $(k+1)\in\{p_{j+1},p_{j+2},\ldots\}$, or (2) $(k+1)\in\{p_1,\ldots,p_j\}$ and $k\in\{p_{j+1},p_{j+2},\ldots\}$.

For any permutation p of N and any positive integer j, we investigate the number of intervals I_k to which j belongs. If the set $\{p_1,\ldots,p_j\}$ is written as the union of the disjoint collection of intervals $[1,b(j)]$, $J(j,2),\ldots,J(j,n_j)$, then we see that $j\in I_k$ in exactly two cases: (a) when

$k \in \{p_1,\ldots,p_j\}$ and k is a right-hand endpoint of one of these intervals, or (b) $(k+1) \in \{p_1,\ldots,p_j\}$ and $(k+1)$ is a left-hand endpoint of some $J(j,i)$. Since the decomposition of the set $\{p_1,\ldots,p_j\}$ into intervals has (n_j-1) gaps, there are exactly $2(n_j-1)+1=2n_j-1$ possible values for k. These considerations show that Theorem 3 is equivalent to Theorem 2.

Theorem 3 requires that any integer j belong to at most $(K-1)$ different intervals I_k, so $2n_j-1 \leqslant K-1$ for all j, and Agnew's criterion in Theorem 2 holds. Obviously, the condition of Theorem 2 implies that of Theorem 3 in view of the preceding discussion.

8. Proof of the Sufficiency of Agnew's Condition. We now prove that a permutation which satisfies the hypothesis of Theorem 2 produces a sum-preserving rearrangement of any convergent series of complex numbers. An examination of the proof shows that it is valid for convergent series in any Banach space (merely use norms instead of absolute values). It should be pointed out that both Levi and Pleasants remark that, because of the combinatorial nature of their proofs, their results hold in more general contexts. Recently, O. Adrian [1] has obtained some sufficient conditions for a permutation to preserve the sum of certain convergent series in Banach spaces, but these results are of a somewhat different character than those discussed here.

Suppose then that p is a permutation of N which satisfies Agnew's condition and that Σu_k is an arbitrary convergent series with sum s. Let $\varepsilon > 0$ be given. By the Cauchy Convergence Condition, there is a positive integer m so that for all $n \geqslant m$ and $q \geqslant 1$,

$$\left| \sum_{k=n+1}^{n+q} u_k \right| < \varepsilon/2M,$$

M being the maximum number of intervals into which each $\{p_1,\ldots,p_j\}$ can be decomposed according to Theorem 2. (We refer to the notation of Section 5 above.) If j is large enough so that $b(j) \geqslant m$, then

$$\sum_{k=1}^{j} u_{p_k} - \sum_{k=1}^{b(j)} u_k = \Sigma_2 u_k + \cdots + \Sigma_{n_j} u_k,$$

where Σ_i denotes summation on all k in the interval $J(j,i)$, $i=2,\ldots,n_j$. Each of the sums on the right is less than $\varepsilon/2M$ in absolute value, so it follows that the absolute value of the left-hand side is less than $\varepsilon/2$. Furthermore, it follows from the Cauchy Convergence Condition that for all $n \geqslant m$,

$$\left| \sum_{k=1}^{n} u_k - s \right| \leqslant \varepsilon/2M \leqslant \varepsilon/2.$$

If j is large enough so that $b(j)$ exceeds m, we have

$$\left| \sum_{k=1}^{j} u_{p_k} - s \right| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so $\Sigma\, u_{p_k}$ also converges to s.

9. Proof of the Necessity of Agnew's Condition. In order to show that the hypothesis of Theorem 2 is a necessary condition for a rearranged series to converge to the same sum as the original series, we follow some of the ideas in [7]. Before getting to the details, we give some preliminary observations which are valid for any permutation p of N.

Recall from Section 5 that for each j one can write $\{p_1,\ldots,p_j\} = [1, b(j)] \cup J(j,2) \cup \cdots \cup J(j,n_j)$. There are (n_j-1) gaps in this decomposition. Since $B(j)$ is the maximum of $\{p_1,\ldots,p_j\}$ and $B(j) \geqslant j$,

$$\sum_{k=1}^{j} u_k - \sum_{k=1}^{j} u_{p_k} = \left(\sum_{k=1}^{B(j)} u_k - \sum_{k=1}^{j} u_{p_k} \right) + \left(\sum_{k=1}^{j} u_k - \sum_{k=1}^{B(j)} u_k \right) = \sum{}' u_k + \sum{}'' u_k,$$

where the indices of summation in Σ' run over the integers in the gaps of the decomposition of $\{p_1,\dots,p_j\}$, and $\Sigma'' u_k$ is either zero if $B(j)=j$ or is equal to

$$\left(-\sum_{k=j+1}^{B(j)} u_k\right)$$

otherwise. If Σu_k converges, the Cauchy Convergence Condition assures that $\Sigma'' u_k$ can be made arbitrarily small in absolute value for all sufficiently large j.

Now, suppose that p is a permutation of N which does not satisfy the hypothesis of Theorem 2. We construct a series which converges to zero, but whose rearrangement by the permutation p does not. Specifically, it is shown for a certain convergent series that the absolute value of the summation, $\Sigma' u_k$, is $+1$ for infinitely many values of j. This proof can be modified to hold in a Banach space by selecting a nonzero element x of the space and replacing the terms $+1, -1, +1/2, -1/2,\dots$ in the construction by $x, -x, \frac{1}{2}x, -\frac{1}{2}x,\dots$. With this replacement, the summation $\Sigma' u_k$ will have norm equal to the positive number $\|x\|$ for infinitely many j.

Since the condition of Theorem 2 is assumed not to hold for the permutation p, for each positive integer i there are infinitely many positive integers j so that $n_j \geqslant i+2$. Let j_1 be such that $n_{j_1} > 3$. Assuming that $j_1,\dots,j_i$ have been defined, let j_{i+1} be such that $n_{j_{i+1}} > (i+1)+2$ *and* $B(j_i) \leqslant b(j_{i+1})$. This latter condition is possible because $b(j)\to\infty$ as $j\to\infty$. The sequence $\{j_i\}$ of positive integers thus defined is strictly increasing. We construct a convergent series Σu_k as follows.

Let $u_1 = 1$ and let $u_k = -1$ when $k = b(j_1)+1$. Set $u_k = 0$ for all other k satisfying $1 \leqslant k \leqslant B(j_1)$. We continue by induction.

If u_k has been defined for all k such that $1 \leqslant k \leqslant B(j_{i-1})$, set $u_k = -1/i$ when k is one less than the left-hand endpoint of each of the intervals $J(j_i, q)$, $2 \leqslant q \leqslant i+2$; set $u_k = 1/i$ when k *is* one of the left-hand endpoints of the same set of intervals; and let u_k equal zero for all other k, $B(j_{i-1})+1 \leqslant k \leqslant B(j_i)$. The resulting series, Σu_k, clearly converges to zero, and the $B(j_i)$th partial sums are equal to zero for each i. On the other hand, $\Sigma' u_k = -1$ for $j = j_1, j_2,\dots$, so it is false that Σu_{p_k} converges to zero.

Although it seems possible, in view of the above proof, that there might be a permutation which preserves the property of convergence of all series but not necessarily convergence to the same sum, it will be shown below that this is not the case. (See also [7].)

10. Summability Considerations. Theorem 2 arose in the context of a matrix transformation of one sequence into another and the consequent requirement that the matrix transforms convergent sequences into convergent sequences. See, for example, [5, p. 43] for these considerations. There is a parallel, almost equivalent theory of matrix transformations of the terms of one series into the terms of another series. In fact, for series with bounded partial sums, the two approaches are equivalent [3, p. 86]. The basic result about "series-to-series" transformations was given by Vermes in [11].

VERMES'S THEOREM. *The infinite matrix (b_{jk}) transforms every convergent series, Σu_k, into a convergent series, Σv_j, where $v_j = \Sigma_k b_{jk} u_k$ if and only if*

(1) *there is an $M > 0$ so that $\Sigma_{k=1}^{\infty} |\Sigma_{j=1}^{n} (b_{jk} - b_{j,k+1})| < M$ for every n, and*

(2) *for each k, $\Sigma_{j=1}^{\infty} b_{jk}$ converges to B_k.*

Moreover, Σv_j converges to $B_1 s + \Sigma_{k=1}^{\infty} (B_k - B_{k+1})(s_k - s)$, where s_k is the kth partial sum of Σu_j and s is the sum of the series Σu_j.

Suppose that p is a permutation of N. A rearrangement of the convergent series Σu_k by the permutation p can be considered as a series-to-series transformation in the following way. Set $b_{jk} = 1$ if $k = p_j$ and $b_{jk} = 0$ if $k \neq p_j$. If $v_j = \Sigma_k b_{jk} u_k$, then $\Sigma v_j = \Sigma u_{p_j}$, the rearrangement produced by p. Since p is one-to-one, each column of (b_{jk}) contains exactly one nonzero term, so $\Sigma_j b_{jk} = 1$ for every k, and condition (2) of the theorem is automatically satisfied with $B_k = 1$.

The first condition of the theorem can be examined in the following way. Observe that since p is an onto mapping, if k is any arbitrary positive integer, then $k=p_s$ and $(k+1)=p_t$ for some positive integers s and t. If n is any integer, then $\Sigma_{j=1}^{n} b_{jk}=1$ if $n \geqslant s$ and is 0 if $1 \leqslant n < s$. Similarly, we have $\Sigma_{j=1}^{n} b_{j,k+1}=1$ if $n \geqslant t$ and is 0 if $1 \leqslant n < t$. Set $f(n,k)=\Sigma_{j=1}^{n}(b_{jk}-b_{j,k+1})$. Then $f(n,k)=+1$ if $s \leqslant n < t$, $f(n,k)=-1$ if $t \leqslant n < s$, and $f(n,k)=0$ for all other values of n. Hence, in the notation of Theorem 3, $|f(n,k)|=1$ if and only if $n \in I_k$. For each n, the series, $\Sigma_k |f(n,k)|$, has only finitely many nonzero terms, and its sum is equal to the number of different intervals I_k to which n belongs. The first condition of Vermes's Theorem will be satisfied exactly when there is an upper bound to the number of intervals I_k to which every positive integer n can belong. This requirement is, of course, the hypothesis of Theorem 3.

We observe that the rearranged series, Σv_j, must converge to the same sum as that of Σu_k, since $B_k=1$ for every k.

Acknowledgment. The author wishes to thank P. A. B. Pleasants for bringing References [4] and [6] to his attention.

References

1. O. Adrian, Asupra unui criteriu privitor la păstrarea convergenţei de către o serie supusă la o permutare într-un spaţiu Banach, Stud. Cerc. Mat., 28 (1976) 693–697.
2. R. P. Agnew, Permutations preserving convergence of series, Proc. Amer. Math. Soc., 6 (1955) 563–564.
3. R. G. Cooke, Infinite Matrices and Sequence Spaces, London, 1950.
4. U. C. Guha, On Levi's theorem on rearrangement of convergent series, Indian J. Math., 9 (1967) 91–93.
5. G. H. Hardy, Divergent Series, Oxford, 1949.
6. F. W. Levi, Rearrangement of convergent series, Duke Math. J., 13 (1946) 579–585.
7. P. A. B. Pleasants, Rearrangements that preserve convergence, J. London Math. Soc. (2), 15 (1977) 134–142.
8. J. H. Smith, Rearrangements of conditionally convergent real series with preassigned cycle type, Proc. Amer. Math. Soc., 47 (1975) 167–170.
9. G. S. Stoller, The convergence-preserving rearrangements of real infinite series, Pacific J. Math., 73 (1977) 227–230.
10. Q. F. Stout, Rearrangements of Convergent Series (to appear).
11. P. Vermes, Series-to-series transformations and analytic continuation by matrix methods, Amer. J. Math., 71 (1949) 541–562.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

ON THE LAW OF LARGE NUMBERS, INFINITE GAMES, AND CATEGORY

C. G. MÉNDEZ

Department of Mathematical Sciences, Metropolitan State College, Denver, CO 80204

In 1941, J. C. Oxtoby and S. M. Ulam [3, p. 877] showed (via a footnote) that the law of large numbers is false in the sense of category, i.e., the set of real numbers of the unit interval such that in their infinite dyadic development the number of ones in the first n places divided by n tends to one-half is of the first category (although of measure one).

Such a set (as illustrated in their remark) can be exhibited as a countable union of nowhere dense sets. On the other hand, a game-theoretical argument could have been provided within the framework of the Banach-Mazur game originally introduced in 1928 [2] where a couple of players take turns alternately by selecting a subinterval of a previously chosen one, starting with the unit interval (that is, a finite sequence of zeros and ones) in an attempt to win by steering the intersection of this sequence of nested intervals into a dealt set A in the case of player I or into the complement of A in the case of player II.

If we were to identify A for the purposes of this discussion with the set of all sequences of zeros and ones for which the limit of the sequence of "averages" is $1/2$, then it can be easily shown that player II has a winning strategy in this game; for he can undo whatever player I does by, let us say, always choosing a sufficiently large sequence of ones. Since Banach and Mazur have shown that there is a winning strategy for player II if and only if A is of first category, we can conclude that the law of large numbers is false in the sense of category ([**2**, p. 28]).

In this article we would like to use similar arguments to show Theorem 1 below. First we have:

Definition 1. Let X be a topological space. A set $A \subset X$ is said to be of first category in X if A is the countable union of nowhere dense subsets of X. A set $B \subset X$ is said to be residual in X if B is the complement of a set of first category in X.

Theorem 1. *The set of all real numbers of the unit interval determined by all sequences of zeros and ones for which there is a subsequence of the sequence of "averages" whose limit is $1/2$ is residual in the unit interval.*

Note that if $I=[0,1]$, $N=$ the set of natural numbers, and if, given $x \in I$, $f_x \in \{0,1\}^N$ were to be the dyadic expansion of x, then the sequence of averages associated with x may be defined by

$$a_x(n) = \frac{1}{n} \sigma_{y<n} f_x(y) \qquad n \in N.$$

Moreover, if $B=\{x\colon a_x$ has a subsequence converging to $\frac{1}{2}\}$, then $B \supset \cap_n B_n$, where $B_n=\{x\colon$ there is an $m>n$ for which $a_x(m)=\frac{1}{2}\}$, and the interior of each B_n is dense in I. Hence, B can be exhibited to be residual in the unit interval.

Interestingly enough, one can also provide a parallel game-theoretic argument as follows:

Proof of Theorem 1. Let (A, B) where $B=\{x\colon a_x$ has a subsequence converging to $\frac{1}{2}\}$, and let $A=I-B$ be a Banach-Mazur game played by choosing blocks of digits of arbitrary lengths in the dyadic development of x.

Clearly, player II has a winning strategy in (A,B) by choosing always the complement (mod 2) of the block just chosen by player I in the preceding move. Hence, by the Banach-Mazur theorem on infinite games ([**2**, p. 28]), we have that B is residual in the unit interval. Q.E.D.

In helping me elucidate these ideas, Professor J. C. Oxtoby has suggested that, if we were to let $B'=\{x\colon \overline{\lim}\, a_x(n)=1$ and $\underline{\lim}\, a_x(n)=0\}$, and $A=I-B'$, then again there is a winning strategy for player II in (A, B'); for he has only to choose (at his successive turns) a sufficiently long block of 1's and 0's, alternately. Hence B' is residual. But if $x \in B'$, then $a_x(n)$ oscillates between values close to 1 and close to 0, and in each such oscillation it has to assume the value $\frac{1}{2}$ at least once. Therefore, $B' \subset \{x\colon a_x(n)=\frac{1}{2}$ for infinitely many $n\}$, and we conclude again that B is residual (as well as that the law of large numbers is false in the sense of category).

The author wishes to thank Professor J. C. Oxtoby for his helpful comments and suggestions, and Professor S. M. Ulam, who directed his doctoral dissertation where some of these results first appeared.

References

1. C. G. Méndez, On Borel and analytic structures and the interrelation of measure and category in an abstract

setting, in On Sigma-Ideals of Sets and on Measures in 2^x, Ph.D. dissertation, University of Colorado, Boulder, 1974.

2. J. C. Oxtoby, Measure and Category, Springer-Verlag, New York, 1971.

3. J. C. Oxtoby and S. M. Ulam, Measure-preserving homeomorphisms and metrical transitivity, Ann. of Math., 42 (2) (1941) 874–920.

A RATIONAL POLYNOMIAL WHOSE GROUP IS THE QUATERNIONS

RICHARD A. DEAN

Department of Mathematics, California Institute of Technology, Pasadena, CA 91125

1. Introduction. In presenting Galois theory most textbooks give examples of polynomials over the field Q of rational numbers whose groups have low order. Few, if any, give an example of an equation whose group is the quaternion group of order 8. This question was investigated in the early 1900's by Gösta Bucht [1] and Franz Mertens [2], [3], who gave methods for constructing such examples. Recently Olga Taussky [4] has discussed this problem in a very elegant and general way. She points out that the question of whether a Galois extension E of a field F is a quaternion extension is related to problems involving sums of two and three squares.

In this note we follow the suggestions of Mertens to show that

$$q(x)=x^8-72x^6+180x^4-144x^2+36$$

has the quaternion group over Q. Our chief prerequisites are these:

1. The fundamental theorem of Galois theory (FTGT), which describes the correspondence between subfields of a splitting field E of a separable polynomial $p(x)=(x-\theta_1)\cdots(x-\theta_n)$ over a field F and the subgroups of the group $\mathcal{G}(E/F)$ of automorphisms of E fixing each element of F. (In this note we shall assume that all fields have characteristic 0.)

2. The role of the discriminant $\Delta^2=\Pi_{i<j}(\theta_i-\theta_j)^2$: The group $\mathcal{G}(E/F)$ is a subgroup of the alternating group $\mathcal{A}_n$ iff Δ is in F.

3. A lemma on quadratic extensions: If $F\neq F(\sqrt{a})$, then $F(\sqrt{a})=F(\sqrt{b})$ iff a/b is a square in F.

4. The quaternion group is the only group of order 8 having three cyclic subgroups of order 4. Every subgroup of the quaternions is normal.

We begin with a search for necessary conditions for a normal extension T of the rationals Q to have the quaternions as its group of automorphisms. The FTGT requires the complete subfield structure of T to be that shown in Fig. 1. Because all subgroups are normal, all subfields are normal extensions of Q. We recognize that $\mathcal{G}(S/Q)$ is the Klein 4-group. The important

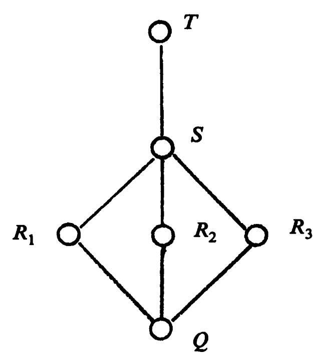


FIG. 1. The subfield structure of a normal extension T of the rationals Q whose Galois group is the quaternions.

observation is that the field T is a *cyclic* extension of degree 4 over each subfield R_i. Thus we need to know something of the structure of cyclic extensions of degree four. The following two theorems are crucial to the construction.

THEOREM 1. *Let T be a cyclic extension of degree 4 over R. Then there exists $d \in R$ such that*
(i) *d is not a square in R,*
and there exist e, f in R such that
(ii) *$T = R(\sqrt{(e + f\sqrt{d})})$ and $d(e^2 - f^2 d)$ is a square in R.*

We should remark that (ii) is equivalent to the condition that d be the sum of two squares in R. The next theorem is the converse of Theorem 1 with a little more information thrown in.

THEOREM 2. *If $d \in R$ is not a square in R and if for $e, f \in R$, $d(e^2 - f^2 d)$ is a square in R, then $T = R(\sqrt{(e + f\sqrt{d})})$ is a cyclic extension of degree 4 over R and $p(x) = x^4 - 2ex^2 + (e^2 - f^2 d)$ is the minimal polynomial of $\sqrt{(e + f\sqrt{d})}$ over R.*

We shall prove these theorems in Section 3 and 4.

2. The Construction. It is easy to describe a candidate for the intermediate field S. Take $S = Q(\sqrt{2}, \sqrt{3})$ so that $R_1 = Q(\sqrt{2})$, $R_2 = Q(\sqrt{3})$ and $R_3 = Q(\sqrt{6})$. The trick is to find a quadratic extension T of S, $T = S(\theta)$ so that θ^2 may be simultaneously expressed in the form

$$\theta^2 = e_i + f_i\sqrt{d_i} \text{ where } \sqrt{d_i} \text{ is not in } R_i \text{ yet } d_i(e_i^2 - f_i^2 d_i) \text{ is a square in } R_i.$$

The articles of Mertens suggest defining $\theta^2 \in S$ by

$$\theta^2 = (2 + \sqrt{2})(2 + \sqrt{3})(3 + \sqrt{6}) = 18 + 12\sqrt{2} + 10\sqrt{3} + 7\sqrt{6}.$$

We may rewrite θ^2 in three ways. Please interpret the next display as a way of defining e_i, f_i, and d_i, for $i = 1, 2$, and 3 in turn.

$$\theta^2 = \begin{cases} (18 + 12\sqrt{2}) + (10 + 7\sqrt{2})\sqrt{3} \\ (18 + 10\sqrt{3}) + (12 + 7\sqrt{3})\sqrt{2} \\ (18 + \ 7\sqrt{6}) + (12 + 5\sqrt{6})\sqrt{2} \end{cases}$$

Now verify directly that, for each i, $d_i(e_i^2 - f_i^2 d_i)$ is a square in R_i:

$$\text{In } R_1 = Q(\sqrt{2}), \quad 3\left[(18 + 12\sqrt{2})^2 - (10 + 7\sqrt{2})^2 \cdot 3\right] = [3(2 + \sqrt{2})]^2.$$

$$\text{In } R_2 = Q(\sqrt{3}), \quad 2\left[(18 + 10\sqrt{3})^2 - (12 + 7\sqrt{3})^2 \cdot 2\right] = [2(3 + 2\sqrt{3})]^2.$$

$$\text{In } R_3 = Q(\sqrt{6}), \quad 2\left[(18 + 7\sqrt{6})^2 - (12 + 5\sqrt{6})^2 \cdot 2\right] = [2(3 + \sqrt{6})]^2.$$

Now using Theorem 2 we conclude that $T = S(\theta)$ is cyclic of degree 4 over each R_i. Hence $[T : Q] = 8$. Our next step is to find a polynomial over Q for which T is its splitting field. When we do, it then follows automatically that T is a normal (Galois) extension of Q and indeed we see that then $\mathcal{G}(T/Q)$ is the quaternions since it has the three cyclic subgroups of order 4 corresponding to the three cyclic extensions T/R_i of degree 4. From Theorem 2 we know that T is the splitting field of

$$p(x) = x^4 - 2e_1x^2 + (e_1^2 - 3f_1^2)$$

over $R_1 = Q(\sqrt{2})$ and $p(\theta) = 0$. We hope to find a conjugate to θ and a conjugate equation by applying the automorphism of S that sends $\sqrt{2}$ into $-\sqrt{2}$ and fixes $\sqrt{3}$. Let $(\bar{\ })$ denote this conjugation so that

$$\bar{e}_1 = 18 - 2\sqrt{2} \quad \text{and} \quad \bar{f}_1 = 10 - 7\sqrt{2}.$$

Define

$$\phi^2=(2-\sqrt{2})(2+\sqrt{3})(3-\sqrt{6})$$

so that ϕ is a root of

$$\bar{p}(x)=x^4-2\bar{e}_1x^2+(\bar{e}_1^2-3\bar{f}_1^2)$$

over $R_1=Q(\sqrt{2})$. Note that $\theta^2\phi^2=6(2+\sqrt{3})^2$ so that $\theta\phi=(2+\sqrt{3})\sqrt{6}$ is in T. Since θ, $\sqrt{6}$, and $(2+\sqrt{3})$ lie in T it follows that ϕ is in T as well.

Using Theorem 2 we find that $\bar{p}(x)$ is irreducible over $Q(\sqrt{2})$. Thus θ and ϕ are roots of

$$q(x)=p(x)\bar{p}(x)=x^8-72x^6+180x^4-144x^2+36,$$

which we claim is the desired polynomial over Q!

Since $T=R_1(\theta)=R_1(\phi)$ is a normal extension of R_1, it follows that $q(x)$ splits in T. On the other hand, θ belongs to no proper subfield of T so that T must be the splitting field of $q(x)$ over Q. (Alternatively one could argue easily that $q(x)$ is irreducible over Q and so $[Q(\theta):Q]=8$.) Thus the construction is complete and its requisite properties have been established. An interesting homework exercise is to find generators of $\mathcal{G}(T/Q)$ as permutations on the roots of $p(x)$.

3. Proof of Theorem 1. Let T be a cyclic extension of degree 4 over R. By the FTGT its subfield structure is $T\supset S\supset R$ where $S=R(\sqrt{d})$; and so d is not a square in R, and $T=S(\theta)$ where $\theta^2=e+f\sqrt{d}$ is in S. We have $\theta^2-e=f\sqrt{d}$, and so θ is a root of

$$p(x)=x^4-2ex^2+(e^2-f^2d).$$

This polynomial is irreducible over R, for surely it does not have a cubic irreducible factor since 3 does not divide $4=[T:R]$. While if it factored as the product of two quadratic polynomials, θ would be the root of one and so belong to S, the only quadratic subfield of T.

The roots of $p(x)$ are θ, $-\theta$, ϕ, and $-\phi$ where $\phi^2=e-f\sqrt{d}$.

The discriminant of $p(x)$ is

$$\Delta^2=(2\theta)^2(2\phi)^2(\theta-\phi)^2(\theta+\phi)^2(-\theta-\phi)^2(-\theta+\phi)^2=16\,\theta^2\phi^2(\theta^2-\phi^2)^4=16^2(e^2-f^2d)f^4d^2.$$

Since $\mathcal{G}(T/R)$ is cyclic of order 4, it is not a subgroup of the alternating group on four letters. This means that Δ is not in R and thus (e^2-f^2d) is not a square in R. To show that $d(e^2-f^2d)$ is a square in R we argue as follows.

Since $T=S(\theta)=S(\phi)$ and $[T:S]=2$, it follows that ϕ/θ is a square *in* S. Now $\phi/\theta=(e-f\sqrt{d})/(e+f\sqrt{d})=(e-f\sqrt{d})^2/(e^2-f^2d)$. We conclude that (e^2-f^2d) is a square in S; say,

$$e^2-f^2d=(r+s\sqrt{d})^2=r^2+s^2d+2rs\sqrt{d}.$$

Thus $2rs=0$. If $s=0$, then $e^2-f^2d=r^2$ is a square in R, contrary to what we have just seen. Therefore, $r=0$, and so $e^2-f^2d=s^2d$. Multiplying through by d we have $d(e^2-f^2d)=s^2d^2$, so that $d(e^2-f^2d)$ is a square in R.

4. Proof of Theorem 2. By hypothesis $d(e^2-f^2d)$ is a square in R and d is not a square in R. Thus (e^2-f^2d) is not a square in R. Moreover $e+f\sqrt{d}$ does not have a square root in $R(\sqrt{d})$. Indeed, if

$$e+f\sqrt{d}=(r+s\sqrt{d})^2=r^2+s^2d+2rs\sqrt{d},$$

then $f=2rs$ and $e=r^2+s^2d$ so that

$$e^2-f^2d=(r^2+s^2d)^2-4r^2s^2d=(r^2-s^2d)^2$$

would be a perfect square in R, contrary to hypothesis. Thus there is a tower of quadratic extensions:

$$R\subset R(\sqrt{d})=S\subset S(\sqrt{(e+f\sqrt{d})}).$$

Now let $\theta^2 = e + f\sqrt{d}$. As in Section 3, θ is a root of

$$p(x) = x^4 - 2ex^2 + (e^2 - f^2 d) = [x^2 - (e + f\sqrt{d})][x^2 - (e - f\sqrt{d})].$$

Since each of the last two factors is irreducible over $R(\sqrt{d})$, it follows that $p(x)$ is irreducible over R. Since

$$(e - f\sqrt{d})/(e + f\sqrt{d}) = (e - f\sqrt{d})^2/(e^2 - f^2 d) = (\sqrt{d})^2(e - f\sqrt{d})^2/d(e^2 - f^2 d)$$

and $d(e^2 - f^2 d)$ is a square in R, it follows that

$$S(\sqrt{(e + f\sqrt{d})}) = S(\sqrt{(e - f\sqrt{d})}).$$

From this it follows that $p(x)$ splits in T but in no proper subfield. So T is the splitting field of $p(x)$ over R and is thus a normal extension of degree 4. Its group is either cyclic or the Klein 4-group. Since only the Klein 4-group is a subgroup of the alternating group on 4 letters, the structure of $\mathcal{G}(T/R)$ can be resolved by examining the discriminant of $p(x)$. As in Section 3 we calculate

$$\Delta^2 = 16^2(e^2 - f^2 d)f^4 d^2$$

and, since $(e^2 - f^2 d)$ is not a square in R, Δ is not in R; hence $\mathcal{G}(T/R)$ is the cyclic group of order 4. An interesting homework exercise is to determine a generator for this group as a permutation on the roots of $p(x)$.

References

1. Gösta Bucht, Arkiv Math. Astron. and Physik, 6. No. 30.
2. Franz Mertens, Akadamie der Wiss., Wien, Sitzb., Abt. IIa 111 (1902).
3. ______, Gleichungen 8-ten Grader mit Quaterniongruppe, Akadamie der Wiss., Wien, Sitzb., Abt. IIa, 125 (1916) 735–740.
4. Olga Taussky, Sums of squares, this MONTHLY, 77 (1970) 805–830.

ON MONTEL'S PROOF OF THE GREAT PICARD THEOREM

DOUGLAS S. BRIDGES

University College, Buckingham MK18 1E9, England

In proving the Great Picard Theorem, several authors use an argument, originally due to Montel [3], in which normal families of analytic functions play a central role. In this approach, Schottky's Theorem [2, Ch. 12, 3.3] is used to prove the local boundedness of a certain family of Picard functions; the normality of this family then follows from Montel's Theorem and entails normality of the family of all Picard functions [2, Ch. 12, 4.1]. This normality is then applied to a well-chosen sequence of Picard functions to produce the Picard result.

The purpose of this note is to point out that, by the introduction of one elementary lemma, all the apparatus of normal families can be removed from this type of argument. This means that, if we follow the presentation of the Bloch and Schottky theorems in [2], we have a proof of the Great Picard Theorem that depends on nothing more sophisticated than Rouché's Theorem and the Maximum Principle.

For $0 \leqslant r < s$, $A(r,s)$ will denote the open annulus $\{z \in C : r < |z| < s\}$, and $\bar{A}(r,s)$ its closure. By a *Picard function* we shall mean a function that is analytic on $A(0,1)$ and omits the values 0, 1 from its range.

LEMMA. *For each $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that: if f is a Picard function, $0 < |\zeta| \leqslant 1/2$, and $|f(\zeta)| \leqslant \varepsilon$, then $|f(z)| \leqslant \delta(\varepsilon)$ whenever $|z| = |\zeta|$.*

Proof. We first consider the case where $|\zeta|=1/2$ and $|f(\zeta)|\leqslant\varepsilon$. Choose a positive integer ν so that $\cos(\pi/\nu)>7/8$, and define $z_k=\zeta\exp(k\pi i/\nu)$ for $k=0,1,\ldots,\nu$. Then

$$\begin{aligned}|z_{k+1}-z_k| &= |\zeta|\cdot|\exp(k\pi i/\nu)|\cdot|\exp(\pi i/\nu)-1|\\ &= 2^{-1}|\cos(\pi/\nu)+i\cdot\sin(\pi/\nu)-1|\\ &= 2^{-1}(2-2\cos(\pi/\nu))^{1/2}.\end{aligned}$$

Hence

$$|z_{k+1}-z_k|^2=2^{-1}(1-\cos(\pi/\nu))<1/16,$$

and so $|z_{k+1}-z_k|<1/4$. We now see that the balls $\bar{B}(z_k,1/4)$ $(k=0,\ldots,\nu)$ form a cover of the semicircle $\{\zeta\exp(i\phi):0\leqslant\phi\leqslant\pi\}$ such that the center of the $(k+1)$th ball belongs to the kth.

Now, according to Schottky's Theorem [2, Ch. 12, 3.3], to each $\alpha>0$ there corresponds $C(\alpha)>0$ such that, if f is analytic and omits the values $0,1$ in a region containing the ball $\bar{B}(a,1/4)$, and if $|f(a)|\leqslant\alpha$, then $|f(z)|\leqslant C(\alpha)$ whenever $|z-a|\leqslant 1/4$. Define positive numbers δ_k as follows: $\delta_0=\varepsilon$ and

$$\delta_{k+1}=\max\nolimits_{n=0,\ldots,k}C(\delta_n)\qquad(k=0,\ldots,\nu-1).$$

If $|f(\zeta)|\leqslant\varepsilon$, then a simple induction argument shows that $|f(z)|\leqslant\delta_k$ whenever $|z-z_k|\leqslant 1/4$ and $k=0,1,\ldots,\nu$. Setting $\delta(\varepsilon)=\delta_\nu$, we now see that $|f(z)|\leqslant\delta(\varepsilon)$ whenever $z=|\zeta|\exp i\phi$ with $0\leqslant\phi\leqslant\pi$. Similar considerations involving the points $|\zeta|\exp(-k\pi i)$ $(k=0,1,\ldots,\nu)$ enable us to show that $|f(z)|\leqslant\delta(\varepsilon)$ for all z with $|z|=1/2$.

If $0<|\zeta|<1/2$ and $|f(\zeta)|\leqslant\varepsilon$, then we only need apply the foregoing to the Picard function $z\rightarrow f(2|\zeta|z)$ to complete the proof. □

We now arrive at:

THE GREAT PICARD THEOREM. *Let f be a Picard function. Then f has either a pole or a removable singularity at* 0.

Proof. For each positive integer n, let

$$\lambda_n=\inf\{|f(z)|:|z|=2^{-n}\},$$

$$\mu_n=\sup\{|f(z)|:|z|=2^{-n}\}.$$

If the sequence (λ_n) is unbounded, then it is increasing: for then, if $\lambda_{n+1}<\lambda_n$ and we choose $m>n+1$ so that $\lambda_m>\lambda_{n+1}$, we have

$$\sup\{|1/f(z)|:2^{-n}\leqslant|z|\leqslant 2^{-m}\}>\sup\{|1/f(z)|:|z|=2^{-n}\text{or}|z|=2^{-m}\},$$

which contradicts the Maximum Principle. In that case, it follows that λ_n increases to ∞, so that $|f(z)|\rightarrow\infty$ as $z\rightarrow 0$ and f has a pole at 0. On the other hand, if (λ_n) is bounded above by some positive number M, then (μ_n) is bounded above by $\delta(M)$ (by our Lemma). Applying the Maximum Principle to f in each of the annuli $\bar{A}(2^{-n-1},2^{-n})$, we then have $|f(z)|\leqslant\delta(M)$ whenever $0<|z|\leqslant 1/2$; it is now easy to prove that f has a removable singularity at 0. □

COROLLARY. *Let f be analytic in $A(0,1)$ with an essential singularity at* 0. *Then f omits at most one complex number from its range.*

The observant reader will have noted that our proof of the Great Picard Theorem is nonconstructive, in that it does not tell us how to compute the (finite) order at 0 of the Picard function f. On the other hand, if f is analytic in $A(0,1)$, has an essential singularity at 0, and omits the value 0 from its range, then the above Corollary tells us that $f(z)$ takes the value 1 but does not enable us to compute a complex number z with $f(z)=1$. It is of interest, therefore, that constructive proofs of both the Great Picard Theorem and its Corollary can be given [1].

References

1. D. S. Bridges, A. Calder, W. Julian, R. Mines, and F. Richman, Picard's Theorem (to appear).

2. J. B. Conway, Functions of One Complex Variable, Graduate Texts in Mathematics, No. 11, 2nd ed., Springer-Verlag, Berlin, 1975.

3. P. Montel, Sur les familles de fonctions analytiques qui admettent des valeurs exceptionnelles dans un domaine, Ann. Ecole Norm. Sup. (3), 23 (1912) 487–535.

MOST MONOTONE FUNCTIONS ARE SINGULAR

TUDOR ZAMFIRESCU

Abteilung Mathematik, Universität Dortmund, Dortmund, West Germany

The history of singular functions, i.e., monotone continuous functions of one real variable with a.e. vanishing derivative, begins in 1904, when H. Lebesgue [1] and H. Minkowski [2] produced their well-known examples. Since then many interesting examples have been published. One of the most recent is due to L. Takács [3], who also presents a good bibliography.

The purpose of this note is to present the following result: *Most monotone functions are singular*. In fact we shall prove the slightly less easy assertion that most functions of uniformly bounded variation have an a.e. vanishing derivative. The monotone case can be treated analogously; therefore a separate proof will not be given. The word *most* has to be understood in the sense of *all, except those in a set of first Baire category*, and will be used only in a space of second Baire category.

Let $\mathcal{C}([0,1])$ be the space of continuous real functions on $[0,1]$ and $\mathcal{V}(V)$ be the family of all functions in $\mathcal{C}([0,1])$ with variation at most $V\in\mathbb{R}$. Clearly, $\mathcal{V}(V)$ is closed in $\mathcal{C}([0,1])$ and therefore is a complete metric subspace with respect to the usual supremum-distance. Thus $\mathcal{V}(V)$ is in itself a space of second Baire category.

Let f_i^- and f_s^- (f_i^+ and f_s^+) be the left (right) lower and upper Dini derivatives of $f:[0,1]\to\mathbb{R}$.

THEOREM 1. *For most functions $f\in\mathcal{V}(V)$, we have, at each point $x\in(0,1]$,*

$$f_i^-(x)=-\infty \quad \text{or} \quad f_i^-(x)\leqslant 0\leqslant f_s^-(x) \quad \text{or} \quad f_s^-(x)=\infty;$$

and, at each point $x\in[0,1)$,

$$f_i^+(x)=-\infty \quad \text{or} \quad f_i^+(x)\leqslant 0\leqslant f_s^+(x) \quad \text{or} \quad f_s^+(x)=\infty.$$

Proof. Let $n\in\mathbb{N}$ and let $\mathcal{V}_n$ be the family of all functions $f\in\mathcal{V}(V)$ such that there exists a point $x\in[0,1]$ for which (i) $x-n^{-1}\geqslant 0$ and

$$n^{-1}\leqslant\left|\frac{f(y)-f(x)}{y-x}\right|\leqslant n \tag{$*$}$$

for each point $y\in(x-n^{-1},x)$, or (ii) $x+n^{-1}\leqslant 1$ and ($*$) holds for each point $y\in(x,x+n^{-1})$.

We show that $\mathcal{V}_n$ is closed in $\mathcal{V}(V)$. Let $f_m\to f$ with $f_m\in\mathcal{V}_n$ and $f\in\mathcal{V}(V)$. There exists $x_m\in[0,1]$ with

$$n^{-1}\leqslant\left|\frac{f_m(y)-f_m(x_m)}{y-x_m}\right|\leqslant n \tag{$**$}$$

for each y in $(x_m-n^{-1},x_m)\subset[0,1]$ or each y in $(x_m,x_m+n^{-1})\subset[0,1]$. We may suppose that x_m converges to some point $x_0\in[0,1]$ and that one of the intervals $(x_m-n^{-1},x_m),(x_m,x_m+n^{-1})$, say the first of these, can be taken for all indices m. (Otherwise consider a subsequence.) Now let $y_0\in(x_0-n^{-1},x_0)$. For m large enough, $y_0\in(x_m-n^{-1},x_m)$ and ($**$) holds with $y=y_0$. Since $f_m\to f$ and $x_m\to x_0$, we get

$$n^{-1} \leqslant \left| \frac{f(y_0)-f(x_0)}{y_0-x_0} \right| \leqslant n,$$

which proves $f \in \mathcal{V}_n$.

Now, we show that $\mathcal{V}(V)-\mathcal{V}_n$ is dense in $\mathcal{V}(V)$. Let $\mathcal{O}$ be an open set in $\mathcal{V}(V)$, choose $f \in \mathcal{O}$, and let $\varepsilon>0$ be such that every function at distance at most ε from f lies in $\mathcal{O}$. Since f is uniformly continuous, there is a $\delta>0$ such that $|x-y|<\delta$ implies $|f(x)-f(y)|<\varepsilon$. We introduce a partition $0=a_0, a_1, \ldots, a_k, a_{k+1}=1$ of $[0,1]$ such that $a_i<a_{i+1}$ and $a_{i+1}-a_i<\delta (i=0,\ldots,k)$. If $f(a_i) \neq f(a_{i+1})$, let $b_i \in (a_i, a_{i+1})$ be such that

$$\left| \frac{f(a_{i+1})-f(a_i)}{b_i-a_i} \right| > n;$$

otherwise let $b_i=a_i$. We construct a continuous function $g:[0,1] \to \mathbb{R}$ in the following way: the restriction of g to $[a_i, a_{i+1}]$ is linear from a_i to b_i and constant on the rest, $g(a_i)=f(a_i)$ and $g(b_i)=g(a_{i+1})=f(a_{i+1})$ $(i=0,\ldots,k)$. Clearly, the distance from f to g does not exceed ε and the variation of g is not greater than that of f. Also, $g \notin \mathcal{V}_n$ by construction. Thus $g \in (\mathcal{V}(V)-\mathcal{V}_n) \cap \mathcal{O}$.

Now, since $\mathcal{V}_n$ is closed and has a dense complement, it is nowhere dense in $\mathcal{V}(V)$.

Let $\mathcal{V}^*$ be the family of those functions $f \in \mathcal{V}(V)$ for which, at some point $x \in (0,1]$,

$$f_i^-(x) > -\infty \quad \text{and} \quad f_s^-(x)<0$$

or

$$f_i^-(x)>0 \quad \text{and} \quad f_s^-(x)<\infty,$$

or, at some point $x \in [0,1)$,

$$f_i^+(x) > -\infty \quad \text{and} \quad f_s^+(x)<0$$

or

$$f_i^+(x)>0 \quad \text{and} \quad f_s^+(x)<\infty.$$

It is easily seen that each $f \in \mathcal{V}^*$ belongs to $\mathcal{V}_n$ for a certain $n \in \mathbb{N}$. We have

$$\mathcal{V}^* = \bigcup_{n=1}^{\infty} \mathcal{V}_n;$$

hence $\mathcal{V}^*$ is of first Baire category and the theorem follows.

Now, if $f \notin \mathcal{V}^*$, then, for every point $x \in [0,1]$, either f is not differentiable at x, or $f'(x)=0$. Since f is differentiable a.e., we have the following:

THEOREM 2. *For most functions $f \in \mathcal{V}(V)$, $f'=0$ a.e..*

With the topology derived from the supremum-metric, we can say nothing similar about the whole space $\mathcal{V}$ of functions of bounded variation from $\mathcal{C}([0,1])$, because

$$\mathcal{V} = \bigcup_{n=1}^{\infty} \mathcal{V}(n)$$

and each $\mathcal{V}(n)$ is nowhere dense in $\mathcal{V}$; hence $\mathcal{V}$ is of the first category itself.

However, we can consider the space $\mathfrak{M}$ of all functions in $\mathcal{C}([0,1])$ that are increasing. Here, without any restriction on the variation, $\mathfrak{M}$ is closed in $\mathcal{C}([0,1])$, hence is of second Baire category in itself. In a similar way as for Theorem 1 we can prove:

THEOREM 3. *For most functions $f \in \mathfrak{M}$, we have, at each point $x \in (0,1]$,*

$$f_i^-(x)=0 \quad \textit{or} \quad f_s^-(x)=\infty,$$

and, at each point $x \in [0,1)$,

$$f_i^+(x)=0 \quad \textit{or} \quad f_s^+(x)=\infty.$$

Since every function $f \in \mathfrak{M}$ is differentiable a.e., the following holds.

THEOREM 4. *For most functions* $f \in \mathfrak{M}$, $f' = 0$ *a.e.*

Analogous phenomena happen for first-order Lipschitz maps. We consider those functions $f \in \mathcal{C}([0,1])$ for which

$$\alpha \leqslant \frac{f(y)-f(x)}{y-x} \leqslant \beta$$

for all pairs of distinct points x, y in $[0,1]$, α and β being fixed real numbers ($\alpha < \beta$). Let $\mathcal{V}_{\alpha,\beta}$ be the family of all such functions; obviously $\mathcal{V}_{\alpha,\beta} \subset \mathcal{V}$ ($\max\{|\alpha|, |\beta|\}$). $\mathcal{V}_{\alpha,\beta}$ is of the second category. We get analogously:

THEOREM 5. *For most functions* $f \in \mathcal{V}_{\alpha,\beta}$, *we have, at each point* $x \in (0,1]$,

$$f_i^-(x) = \alpha \quad or \quad f_s^-(x) = \beta,$$

and, at each point $x \in [0,1)$,

$$f_i^+(x) = \alpha \quad or \quad f_s^+(x) = \beta.$$

THEOREM 6. *For most functions* $f \in \mathcal{V}_{\alpha,\beta}$, *the set*

$$f'^{-1}(\alpha) \cup f'^{-1}(\beta)$$

has measure 1.

References

1. H. Lebesgue, Leçons sur l'intégration et la recherche des fonctions primitives, Gauthier-Villars, Paris, 1904.

2. H. Minkowski, Zur Geometrie der Zahlen, Verhandl. III Intern. Math.-Kongresses Heidelberg (1904) 164–173.

3. L. Takács, An increasing continuous singular function, this MONTHLY, 85 (1978) 35–37.

AN ALGORITHM-INSPIRED PROOF OF THE SPECTRAL THEOREM IN E^n

HERBERT S. WILF

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104

THEOREM. *If A is a real symmetric matrix, there is a real orthogonal matrix Q such that Q^TAQ is diagonal.*

Of course, this is the spectral theorem. It implies that the eigenvalues are real, that there is a pairwise orthogonal complete set of eigenvectors—namely, the columns of Q—and that the dimension of an eigenspace is equal to the algebraic multiplicity of the eigenvalue.

Many proofs grapple with the question of finding enough independent eigenvectors for a multiple eigenvalue, usually one at a time. We shall find the whole matrix Q at once by using the main idea of Jacobi's numerical method for calculating the eigenvalues and vectors, together with a little compactness.

For an $n \times n$ real matrix A we shall use $\mathrm{Od}(A)$ for the sum of the squares of the off-diagonal elements of A, and $O(n)$ will denote the set (group) of $n \times n$ orthogonal matrices.

Suppose we can prove the following.

LEMMA. *If A is a nondiagonal real symmetric matrix, then there is a real orthogonal matrix J such that* $\mathrm{Od}(J^TAJ) < \mathrm{Od}(A)$.

Then the theorem would follow quickly, for let A be real and symmetric. Consider the mapping f that sends an orthogonal matrix P into $f(P) = P^TAP$. For fixed A this is a continuous mapping of $O(n)$, a compact set, and so $f(O(n))$ is compact. Let $f(Q) = D$ be a point at which the continuous function Od attains its minimum value on the image set of f. This value must be

zero, else D could play the role of A in the above lemma, and so it would not minimize Od, concluding the proof.

It remains to prove the lemma. This was done by Jacobi, in his celebrated method of plane rotations for computing eigenvalues and eigenvectors, as follows: Suppose $a_{pq} \neq 0$, $p \neq q$. Then take for J the matrix that agrees with the $n \times n$ identity matrix except in the four positions $(p,p),(p,q),(q,p),(q,q)$, where the entries are $\cos\theta$, $\sin\theta$, $-\sin\theta, \cos\theta$, and the real angle θ is chosen to make $(J^T AJ)_{pq} = 0$. It is easy to check that $\mathrm{Od}(J^T AJ) = \mathrm{Od}(A) - 2a_{pq}^2$, and we are finished.

The proof readily generalizes to the complex Hermitian case.

Research for this paper was supported by the National Science Foundation.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4.*

ARE π, e, AND $\sqrt{2}$ EQUALLY DIFFICULT TO COMPUTE?

L. BAXTER

Department of Computer Science, York University, Downsview, Ontario, Canada

The general problem is to find efficient methods for calculating fundamental mathematical constants. The simplest of methods is chosen, namely, iterative sequences expressed using the operations of $+$, $-$, $\times$, $\div$, and $\sqrt{}$. Whereas iterative sequences exist which show that the time to compute n digits of $\sqrt{2}$ and π are essentially the same, $O(n \cdot \log n \cdot \log\log n)$, no equally efficient method of computing e is known.

The efficient computation of $\sqrt{2}$ and π compared to that of e depends on two factors: rate of convergence and the complexity of multiplication. The iterative sequences converge quadratically, i.e., the number of significant digits doubles with each iteration. Consequently the time to compute n digits of the constant is essentially the time to perform the last iteration. Since this involves multiplications, the Schönhage-Strassen algorithm is used which multiplies two n digit numbers in $O(n \cdot \log n \cdot \log\log n)$ time [**1**, Section 7.5]. (Note that division or extraction of square roots also has this complexity.) If either the convergence is linear or the classical $O(n^2)$ multiplication algorithm is used, then the time to compute n digits is at least $O(n^2)$, which theoretically is as bad as summing the appropriate series.

Specifically, Newton's sequence $x_i = \frac{1}{2}(x_{i-1} + 2/x_{i-1})$ or the more convenient $x_i = x_{i-1}(6 - x_{i-1}^2)/4$ converge quadratically to $\sqrt{2}$. These methods have essentially the same complexity as that based on continued fractions [**2**]. Vieta's formula for π,

$$\frac{2}{\pi} = \sqrt{\tfrac{1}{2}} \cdot \sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\right)} \cdot \sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{1}{2}}\right)}\right)} \cdots,$$

can be defined by the two iterative sequences

$$x_i = \sqrt{\left(\tfrac{1}{2} + \tfrac{1}{2}x_{i-1}\right)}; \qquad y_i = y_{i-1}x_i, \qquad x_1 = 0, \quad y_1 = 1,$$

or by only one

$$y_i = y_{i-1}\sqrt{\left\{\tfrac{1}{2}\left(\frac{y_{i-1}}{y_{i-2}}+1\right)\right\}}; \qquad y_1=1, \quad y_2=\sqrt{\tfrac{1}{2}}.$$

This can be transformed into a sequence which converges to π,

$$z_i = \frac{z_{i-1}}{\sqrt{\left\{\tfrac{1}{2}\left(\frac{z_{i-2}}{z_{i-1}}+1\right)\right\}}}; \qquad z_1=0, \quad z_2=2.$$

Unfortunately convergence is not quadratic. However, there is a quadratically converging method for computing π based on the arithmetic-geometric sequences $x_i=\frac{1}{2}(x_{i-1}+y_{i-1})$, $y_i=\sqrt{(x_{i-1}y_{i-1})}$ ([3]). Note that this pair of sequences is equivalent to the sequence $x_i=\frac{1}{2}[x_{i-1}+\sqrt{\{x_{i-2}(2x_{i-1}-x_{i-2})\}}]$.

In the case of e the familiar $\Sigma 1/i!$ can be converted into the three-step sequence

$$x_i = x_{i-1} + \frac{(x_{i-1}-x_{i-2})^2}{x_{i-1}-x_{i-3}}; \qquad x_1=1, \quad x_2=2, \quad x_3=\frac{5}{2},$$

but this converges too slowly. Now one-step sequences $x_i=F(x_{i-1})$ can converge only to algebraic numbers when F is expressed using $+$, $-$, $\times$, $\div$, and $\sqrt{}$; so the problem is to find a two-step formula $x_i=F(x_{i-1},x_{i-2})$ which converges to e, at least quadratically. Since e is transcendental, it is necessary that $F(x,x)\equiv x$.

Reversing the problem is to ask whether some sequence converges quadratically to some interesting constant; e.g., what do $x_i=\sqrt{\{x_{i-1}^2+a(x_{i-1}-x_{i-2})^2\}}$ or $x_i=x_{i-1}+a(x_{i-1}-x_{i-2})^2$ converge to? Do they involve e?

References

1. A. Aho, J. Hopcroft, and J. Ullman, The Design and Analysis of Algorithms, Addison-Wesley, 1974.
2. J. Dutka, The square root of 2 to 1,000,000 decimals, Math. Comput., 25 (1971), no. 16, 927–930.
3. E. Salamin, Computation of π using arithmetic-geometric mean, Math. Comput., 30 (1976), no. 135, 565–570.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis MO 63121.

EVERY POWER SERIES IS A TAYLOR SERIES

MARK D. MEYERSON

Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402

For every function f, defined and with all its derivatives defined at the origin, we define its Taylor series at the origin as $y=a_0+a_1x+a_2x^2+\cdots$, where $a_n=f^{(n)}(0)/n!$. (Then one studies convergence. The standard example $y=\exp(-1/x^2)$ shows that the Taylor series may fail to converge to the original function.) The following question naturally arises: Is a given power series the Taylor series for some function? We show the answer is yes. Our construction follows that of Gelbaum and Olmsted ([2, pp. 69–70]) who show that $\Sigma_{n=0}^{\infty} n!x^n$ is a Taylor series. My thanks to John Kalme for showing me that example.

THEOREM. *Given any sequence $\{a_n\}_{n=0}^{\infty}$ of real numbers, there is a function with Taylor series $\sum_{n=0}^{\infty} a_n x^n$.*

Proof. For $n=0,1,2,\ldots$, let the function g_n be defined by

$$g_n(x)=\begin{cases} a_n n! & \text{if} \quad |x|\leqslant 1/(2|a_n|n!+1) \\ 0 & \text{if} \quad |x|\geqslant 2/(2|a_n|n!+1) \end{cases}$$

and elsewhere as a "smoothing function" which is monotonic in each interval where it is defined and which makes all derivatives of g_n exist (see [2, p. 40]). Let $f_0=g_0$ and for $n\geqslant 1$ let

$$f_n(x_n)=\int_0^{x_n}\int_0^{x_{n-1}}\cdots\int_0^{x_2}\int_0^{x_1} g_n(x_0)\,dx_0\,dx_1\cdots dx_{n-2}\,dx_{n-1}.$$

Then for $n\geqslant 1$, $|f_n^{(n-1)}(x)|=|\int_0^x g_n(t)\,dt|\leqslant 1$; so integrating $n-k-1$ times we see that $|f_n^{(k)}(x)|\leqslant |x|^{n-k-1}/(n-k-1)!$, where $0\leqslant k\leqslant n-1$ and $f_n^{(0)}=f_n$. By the Weierstrass M-test, $\sum_{n=0}^{\infty} f_n^{(k)}(x)$ converges uniformly on every bounded interval ($k\geqslant 0$). Hence $\sum_{n=0}^{\infty} f_n(x)$ converges to some $f(x)$ and $f^{(k)}(x)=\sum_{n=0}^{\infty} f_n^{(k)}(x)$ for $k\geqslant 1$ (see [6, p. 140]). But $f_n^{(k)}(0)=\delta_{nk}a_n n!$, so $f^{(n)}(0)=a_n n!$ $(n,k\geqslant 0)$. $\square$

COROLLARY. *There are uncountably many functions with any given Taylor series.*

Proof. For f as above and λ real, consider the function h defined by

$$h(x)=f(x)+\lambda\exp(-1/x^2)$$

for $x\neq 0$ and $h(0)=f(0)$. $\square$

Note that we could have centered the Taylor series at any point, not necessarily the origin.

There are many earlier proofs of this result. It seems to have been proved first by E. Borel in 1895 ([1, see p. 44]). He made careful use of the harmonic series to choose coefficients for a power series defining a function with derivatives close to the desired ones. The error is small enough to be accounted for by an analytic function, and so the sum of these two functions is the desired function. Then in 1938 G. Pólya ([4, see p. 244]) gave a proof using machinery he developed for handling systems of equations with infinitely many unknowns. More recently A. Rosenthal ([5]) gave a proof expressing the desired function as an infinite sum of monomials with explicitly calculated exponential coefficients. And closest to the method here is the proof of H. Mirkil ([3]), which is done in n dimensions.

This work was supported by the Naval Academy Research Council Fund.

References

1. Émile Borel, Sur quelques points de la théorie des fonctions, Ann. Sci. École Norm. Sup., (3), no. 12 (1895) 9–55.

2. Bernard R. Gelbaum and John M. H. Olmsted, Counterexamples in Analysis, Holden-Day, 1964.

3. H. Mirkil, Differentiable functions, formal power series, and moments, Proc. Amer. Math. Soc., 7 (1956) 650–652.

4. G. Pólya, Eine einfache, mit funktionentheoretischen Aufgaben verknüpfte, hinreichende Bedingung für die Auflösbarkeit eines Systems unendlich vieler linearer Gleichungen, Comment. Math. Helv., 11 (1938–1939) 234–252.

5. Arthur Rosenthal, On functions with infinitely many derivatives, Proc. Amer. Math. Soc., 4 (1953) 600–602.

6. Walter Rudin, Principles of Mathematical Analysis, 2nd ed., McGraw-Hill, New York, 1964.

Answers to the questions on p. 5.

1. Goethe, *Maximen und Reflexionen*, no. 710 (1829).
2. Hermann Weyl, this MONTHLY, 58 (1951) 523.

SUBMODULES OF FREE MODULES

JOHN ISBELL

Department of Mathematics, State University of New York, Buffalo, NY 14214

In many introductory abstract algebra courses one proves that submodules of finitely generated free modules over a principal ideal domain are free and need no more generators. (Of course one may wish to dig deeper and find a basis displaying the invariants. Even so, most students can use a rest at the point indicated.) Less often, never for all of me, does one want the theorem that submodules of finitely generated free modules over a Dedekind domain are projective, in such a course. Anyway, these two theorems have a proof, practically the same, simpler than any I have seen. It requires introducing the concept of projective module and proving that free modules, and finite direct sums of projective modules, are projective. I guess this may be an instructional gain or an annoyance, depending on the spirit of the course. Also, one appears to get not two but four theorems from one proof (and for general rings I think the four must be different). The proof that for (commutative) integral domains the four reduce to two involves a couple of exercises.

This "one proof" appears, proving another similar result, in P. M. Cohn's *Free Rings and Their Relations*, Academic Press, London-New York, 1971 (proof of Theorem 1.1). I am indebted to G. M. Bergman for the reference.

For $i=1,2,3,4$, for any n, let $P_n^i(A)$ be the proposition: *Every submodule of a free A-module on n generators is* (1) *free,* (2) *free, on no more than n generators,* (3) *projective,* (4) *projective, with n or fewer generators. All $P_0^i(A)$ are trivially true.*

THEOREM. *$P_1^i(A)$ implies $P_n^i(A)$ for all n.*

Proof. It suffices to deduce $P_{k+1}^i(A)$ from $P_1^i(A)$ and $P_k^i(A)$. So in $A^{k+1}=A\oplus A^k$, let S be a submodule. Coordinate projection $f\colon A\oplus A^k\to A^k$ takes S to an image S'; since S' is a submodule of A^k it is (i), and in particular, S' is projective. Thus $f\colon S\to S'$ has a section $s\colon S'\to S$, and S is isomorphic with the direct sum of S' and the kernel S''; $S''=S\cap A$ in $A\oplus A^k$, so $S''\subset A$ is (i), and $S\cong S'\oplus S''$ is also (i).

For an integral domain A, $P_1^2(A)$ or $P_1^4(A)$ implies that every ideal is principal, and conversely; so P_n^2, P_n^4 are equivalent here. A little more thought is needed for:

$$P_1^1 \textit{ implies } P_1^2 \textit{ for commutative rings.}$$

It is just that two elements r,s of a commutative ring A (considered as an A-module) cannot be linearly independent: $sr-rs=0$.

ACKNOWLEDGMENT

The editors thank the following individuals who have refereed manuscripts between mid-1979 and mid-1980: Alexander Abian, J. D. Aczél, R. A. Adams, Jeanne L. Agnew, L. V. Ahlfors, H. L. Alder, G. L. Alexanderson, F. J. Almgren, Jr., Waleed Al-Salam, J. T. Anderson, P. G. Anderson, T. W. Anderson, Jr., G. E. Andrews, E. F. Assmus, Charles Aull, Louis Auslander, Christine Ayoub, A. Baernstein, C. S. Ballantine, S. B. Bank, E. J. Barbeau, P. G. Barker, David Barnette, R. G. Bartle, David Barton, Grace E. Bates, E. F. Beckenbach, P. E. Bedient, G. A. Beer, C. M. Bender, D. C. Benson, H. L. Bentley, S. K. Berberian, E. R. Berkson, E. R. Berlekamp, Elizabeth A. Berman, Gerald Berman, Bruce Berndt, Dorothy L. Bernstein, N. L. Biggs, G. G. Bilodeau, D. W. Blackett, D. K. Blevins, S. L. Bloom, Peter Bloomfield, Ray

Bobo, W. M. Boothby, M. F. Bott, W. E. Boyce, G. U. Brauer, Marilyn Breen, J. E. Brennan, D. W. Brewer, J. P. Brezin, John Brillhart, Demetrios Brizolis, P. L. Brockett, Ezra Brown, J. W. Brown, R. Brown, R. A. Brualdi, A. M. Bruckner, R. C. Buck, P. S. Bullen, R. B. Burckel, R. L. Burden, Donald Burkholder, C. W. Burrill, D. M. Burton, D. W. Bushaw, R. H. Cameron, Alan Candiotti, F. B. Cannonito, J. C. Cantwell, G. T. Cargo, Leonard Carlitz, B. C. Carlson, Stephan Carlson, A. S. Cavaretta, Jr., S. Chakravarthy, Paul Chernoff, R. L. O. Cignoli, D. Clarkson, Amy Cohen, L. W. Cohen, Harvey Cohn, J. H. E. Cohn, Edwin Connell, P. F. Conrad, S. D. Conte, J. B. Conway, C. M. Cordes, Antonio Córdoba, L. J. Corwin, Eugene Couch, J. R. Cowles, V. F. Cowling, H. S. M. Coxeter, Helen F. Cullen, L. J. Cummings, T. W. Cusick, D. A. Darling, Chandler Davis, P. J. Davis, M. M. Day, N. G. de Bruijn, Hans DeBrunner, J. B. Dennin, R. K. Dennis, D. W. DeTemple, Allen Devinatz, C. L. Devito, R. A. Devore, R. J. DiPerna, J. L. Doob, David Drasin, Lester Dubins, Roy Dubisch, James Dugundji, R. A. Duke, P. L. Duren, Meyer Dwass, A. G. Earnest, W. F. Eberlein, G. A. Edgar, H. M. Edwards, Gertrude Ehrlich, J. A. Eidswick, Samuel Eilenberg, Murray Eisenberg, P. D. T. A. Elliott, Erwin Engeler, E. E. Enochs, P. Erdős, D. R. Estes, G. J. Etgen, D. J. Eustice, Siemion Fajtlowicz, J. E. Falk, Ky Fan, K. B. Farmer, G. F. Feeman, M. B. Feldman, A. M. Fink, R. E. Fischer, Harley Flanders, W. H. Fleming, Dominique Foata, D. J. Foulis, J. J. F. Fournier, Evelyn Frank, R. B. Frary, Robert Fray, R. C. Freiwald, Avner Friedman, H. M. Friedman, W. H. Fuchs, Janos Galambos, Steven Galovich, Fred Galvin, D. J. H. Garling, J. B. Garner, Paul Garratt, George Gasper, Jr., B. L. Gelbaum, G. T. Georgantas, Richard Gibbs, Leonard Gillman, Robert Gilmer, Herman Gluck, Abraham Goetz, K. P. Goldberg, R. R. Goldberg, Samuel Goldberg, Seymour Goldberg, J. K. Goldhaber, Jay Goldman, D. L. Goldsmith, J. A. Goldstein, K. R. Goodearl, J. E. Goodman, H. W. Gould, Judith Grabiner, R. L. Graham, Lawrence Gray, Curtis Greene, Gebhard Greiter, E. Griffin, H. C. Griffith, Emil Grosswald, Branko Grünbaum, A. P. Guinand, R. F. Gundy, R. Guralnick, Marjorie G. Hahn, Theodore Hailperin, Wolfgang Haken, J. E. Hall, Paul Halmos, Israel Halperin, R. H. Hanks, Jun-Ichi Hano, Bruno Harris, J. R. Hattemer, Louise Hay, Nickolas Heerema, Hans Heinig, Maurice Heins, Pavol Hell, D. W. Henderson, Leon Henkin, Peter Henrici, I. N. Herstein, Edwin Hewitt, P. J. Hilton, J. G. Hocking, John Hoelzer, R. V. Hogg, V. E. Hoggatt, Jr., A. S. B. Holland, Charles Holmes, R. A. Honsberger, G. C. C. Hsiao, B. W. Huff, F. N. Huggins, D. R. Hughes, Rhonda J. Hughes, T. E. Hull, R. A. Hunt, Mourad Ismail, Martin Jacobsen, Arno Jaeger, R. C. James, J. A. Jenkins, D. R. Jensen, Meyer Jerison, M. A. Jodeit, Jr., Fritz John, C. R. Johnson, W. B. Jones, James Jordan, J. E. Joseph, R. K. Juberg, Erwin Just, Victor Kac, Peter Kall, M. J. Kallaher, J. S. Kalme, Wilfred Kaplan, Irving Kaplansky, L. A. Karlovitz, Jerome Katz, R. P. Kaufman, Nicholas Kazarinoff, A. D. Keedwell, R. B. Kellogg, J. H. B. Kemperman, Irving Kessler, Eleanor Killam, J. O. Kiltinen, Bruce King, A. M. Kirch, W. A. Kirk, Peter Kiss, Gabriel Klambauer, Murray Klamkin, V. L. Klee, Jr., Erwin Kleinfeld, D. J. Kleitman, T. H. Koornwinder, E. F. Krause, H. F. Kreimer, H. F. Ladik, Serge Lang, D. C. Lantz, R. E. Larson, A. C. Lazer, J. D. Leadley, Christopher Lech, D. H. Lehmer, Emma Lehmer, Walter Leighton, P. J. Lelong, Anthony Leung, M. L. Levitan, Lawrence Levy, D. J. Lewis, P. W. Lewis, T. M. Liggett, Charlotte Lin, A. O. Lindstrum, Jr., Marilynn Livingston, J. D. Logan, Judith Q. Longyear, L. H. Loomis, Lee Lorch, George Lorentz, W. F. Lucas, D. H. Luecking, T. H. MacGregor, L. A. Machtinger, B. L. Madison, A. R. Magid, Jeffrey Mandula, O. L. Mangasarian, H. B. Mann, W. R. Mann, Morris Marden, A. W. Marshall, C. W. Marshall, G. E. Martin, L. C. Martins, P. R. Masani, J. M. Masley, J. H. Mason, D. W. Matula, J. G. Mauldon, C. G. McCrory, R. H. McDowell, Amram Meir, Nathan Mendelsohn, Michel Mendès-France, C. G. Méndez, G. H. Meyer, Ernest Michael, R. R. Middlemiss, Angelo Miele, R. J. Mihalek, G. Mill, E. C. Milner, Leon Mirsky, Katherine L. Monti, R. T. Moore, J. M. Moser, Nick Mousouris, Benjamin Muckenhoupt, D. H. Mugler, Francis Murray, Malcolm Murrill, J. B. Muskat, Alex Nagel, M. B. Nathanson, Edward Nelson, Michael Neumann, Marcel Neuts, C.H. Nevison, Albert Nijenhuis, Louis Nirenberg, Ivan Niven, W. P. Novinger, R. D. Nussbaum,

A. M. Odlyzko, Ingram Olkin, F. W. J. Olver, Warren Page, R. S. Palais, F. D. Parker, C. J. Parry, T. D. Parsons, Emanuel Parzen, Donald Passman, Jean J. Pedersen, R. N. Pederson, Daniel Pedoe, G. B. Pedrick, Anthony Peressini, W. J. Pervin, A. C. Peterson, R. S. Phillips, P.F. Pickel, Donald Piele, David Plaisted, B. H. Pourciau, M. W. Pownall, G. B. Preston, H. W. Propper, M. H. Protter, H.-W. Pu, W. R. Pulleyblank, Min-Ping Qian, H. Radjavi, J. B. Rauch, Ludwig Reich, Irving Reiner, H. L. Resnikoff, J. I. Richards, D. G. Rider, F. S. Roberts, R. M. Robinson, Richard Rochberg, C. A. Rogers, V. K. Rohatgi, N. J. Rose, D. P. Roselle, M. A. Rosenlicht, James Ross, C. C. Rousseau, H. L. Royden, L. A. Rubel, Mary Ellen Rudin, Walter Rudin, R. D. Russell, P. J. Ryan, D. G. Saari, T. L. Saaty, G. S. Sacerdote, A. J. Sachs, E. B. Saff, Hans Samelson, J. W. Sanders, D. H. Saracino, D. E. Sarason, C. T. Scarborough, A. T. Schafer, J. J. Schäffer, Doris W. Schattschneider, Martin Schechter, Albert Schild, Hans Schneider, G. E. Schober, I. J. Schoenberg, A. H. Schoenfeld, Steven Schot, Binyamin Schwarz, Abraham Seidenberg, T. I. Seidman, J. L. Selfridge, D. I. Shelupsky, G. J. Sherman, Oved Shisha, Robert Silverman, Martin Silverstein, E. M. Silvia, David Singmaster, J. R. Smart, M. R. Smith, E. Snapper, R. L. Snider, D. C. Soda, Anita E. Solow, E. H. Spanier, Joel Spencer, J. H. Staib, R. P. Stanley, Michael Starbird, Thomas Starbird, Michael Steele, Alan Stein, Sherman Stein, Andrew Sterrett, B. M. Stewart, K. Stolarsky, R. R. Stoll, Gabriel Stolzenberg, W. G. Strang, E. G. Straus, D. W. Stroock, Michio Suzuki, V. Szebehely, L. F. Takács, Olga Taussky-Todd, J. L. Taylor, David Tepper, G. L. Thompson, M. D. Thompson, Wolfgang Thron, Aimo Tietäväinen, W. F. Trench, Carol Tretkoff, Marvin Tretkoff, Selden Trimble, A. C. Tucker, C. T. Tucker, T. W. Tucker, J. J. Uhl, W. R. Utz, Jr., J. D. Vaaler, Charles Vanden Eynden, A. J. van der Poorten, R. Varga, H. V. Waldinger, M. F. Walker, L. E. Ward, Jr., R. H. Warren, Daniel Waterman, M. E. Watkins, H. Wegner, A. E. Weinstein, G. L. Weiss, Allen Weitsman, Robert Welland, J. G. Wendel, J. E. Wetzel, R. L. Wheeden, D. E. White, P. A. White, D. V. Widder, Harold Widom, Sylvia Wiegand, Albert Wilansky, R. L. Wilder, H. S. Wilf, Frederick Wilke, Stephen Willard, L. B. Williams, J. C. S. Wong, D. R. Woodall, Kenneth Wooster, J. W. Wrench, Jr., David Wright, E. M. Wright, Ann Yasuhara, Peter Yff, Cleon Yohe, G. S. Young, Wo-Sang Young, Lawrence Zalcman, David Zeitlin, Martha Zelinka, Daniel Zelinsky, Joseph Zimmer, Jr., A. D. Ziebur, A. C. Zitronenbaum, Martin Zuckerman.

MATHEMATICAL EDUCATION

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

UNIVERSITY MATHEMATICS PLACEMENT TESTING FOR HIGH SCHOOL JUNIORS

AL ADCOCK
Westland High School, Galloway, OH 43119

JOAN R. LEITZEL
Department of Mathematics, The Ohio State University, Columbus, OH 43210

BERT K. WAITS
Department of Mathematics, The Ohio State University, Columbus OH 43210

Over the past fourteen years The Ohio State University, like most other open admissions universities, has observed a sharp decrease in the mathematics skills of entering freshmen.

During this period the OSU mathematics placement examinations and placement algorithms have not changed. The procedure assigns a level of 1, 2, 3, 4, or 5 to each new student; levels 4 and 5 indicate little or no skill in high school algebra and hence a need for remedial work before required university mathematics courses are attempted. Students in level 5 are placed in the first course of the remedial sequence, which begins with topics in arithmetic and includes some elementary algebra. Table A gives the percentage of entering students in levels 4 and 5 since 1966 and shows clearly the increase in the percentage of level 5 students during the early seventies.

TABLE A

Percentage of OSU Autumn Quarter Freshmen
with Remedial Mathematics Placement (Levels 4 and 5)

	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Level 4	15	16	17	16	17	16	16	16	16	16	14	15	16	16
Level 5	13	16	18	17	17	21	23	24	25	26	27	27	26	27

In January, 1978, the mathematics department chairman and the junior class counselor at Westland High School near Columbus met with members of The Ohio State University Department of Mathematics to seek ways of assisting the secondary schools to strengthen the mathematics background of prospective university students. Although many reasons were suggested for the decline in test scores, it was agreed that the large number of students who take no mathematics after their sophomore or junior year probably contributed significantly to the university's remediation problem.

The mathematics placement procedure at Ohio State includes an examination designed by the department. It was suggested that if, during their junior year, students could compare their mathematical skills with university requirements, they might want to take appropriate mathematics courses the next year. Ohio State offered to provide equivalent forms of its placement examination for testing Westland High School juniors. In February, 1978, 353 students were tested. The university machine-scored the tests and provided Westland with the students' scores, with lists of the mathematics courses needed in each intended major program, and with a description of the remedial courses that would be necessary in each case if mathematics skills remained at the tested level.

The Westland teachers and counselors used the test results to recommend to each junior student a specific mathematics course for the senior year. Fortunately, Westland had available a senior review course similar to the university's remedial course. The enrollment in that course, however, had not been large. The review course together with Algebra II and the precalculus course (Advanced Math) provided appropriate courses for all students. The spring scheduling showed dramatic increases in anticipated mathematics enrollments for the senior year, a 73% increase over the previous year. Although not all of these students actually took mathematics courses when classes started the next autumn, school personnel were very pleased with the actual enrollment increases.

Although the increase in senior enrollments indicated a positive response to the testing, changes in placement level as a consequence of taking more high school mathematics could not be checked until autumn, 1979, the quarter in which many of these students were to enter Ohio State. (A summary of these results is given at the end of this article.) Even so, high school and university personnel were sufficiently convinced of the potential of the early testing program to plan for a second year. Other Columbus area high schools, hearing informally of the Westland experiment, asked for similar opportunities. During the academic year 1978–79 the Department of Mathematics extended the program to seven Columbus area high schools. Within this period Gunther Lang, a research associate in the College of Mathematical and Physical Sciences, completed the important task of developing computer programs to give all students an indi-

vidualized description of what their mathematics placement level meant in terms of university requirements for their intended major. Each student could now receive a computer printout sheet indicating the student's test score, the projected Ohio State mathematics placement level, a list of necessary remedial courses (if any), and a list of the university-level mathematics courses required for the student's choice of academic major. All students were encouraged to schedule an appropriate mathematics course in the senior year. These computer programs also provided the high school with a description of the performance of its students grouped according to course preparation and other background information.

The data from testing 1005 students in the 1979 early placement testing program are summarized in Table B. It should be kept in mind that the testing was done in January and February (well before the end of the school year) and that schools made individual decisions concerning which students would be tested. In several schools all students who planned to attend 2-year or 4-year colleges or who were still undecided about college were required to take the examination.

TABLE B
Projected Math Placement Level (MPL) by Course of Enrollment,
Test Given to High School *Juniors*
Winter 1979

Placement Explanation

MPL 4 and 5 are remedial levels indicating little or no skill in elementary and intermediate high school algebra.

MPL 3 is a nonremedial level for all majors except those requiring the calculus courses for science and engineering majors.

MPL 2 is a nonremedial level indicating students are prepared for Ohio State's college algebra/trigonometry courses.

General Math (10.7%)	Algebra I (6.0%)	Geometry (15.7%)
MPL 3: 3 (2.8%)	MPL 3: 5 (8.3%)	MPL 3: 25 (15.8%)
MPL 4: 10 (9.3%)	MPL 4: 2 (3.3%)	MPL 4: 22 (13.9%)
MPL 5: 95 (88.0%)	MPL 5: 53 (88.3%)	MPL 5: 111 (70.3%)
108	60	158

Algebra II (24.7%)	"Senior" Math (3.8%)	No Math (39.1%)
MPL 2: 19 (7.7%)	MPL 2: 4 (10.5%)	MPL 2: 0 (0.0%)
MPL 3: 121 (48.8%)	MPL 3: 19 (50.0%)	MPL 3: 10 (2.5%)
MPL 4: 42 (16.9%)	MPL 4: 13 (34.2%)	MPL 4: 21 (5.3%)
MPL 5: 66 (26.6%)	MPL 5: 2 (5.3%)	MPL 5: 362 (92.1%)
248	38	393

Total Tested: 1005

It has been encouraging to observe that almost 60% of the students taking Algebra II placed at nonremedial levels (MPL 3 or higher). The results would undoubtedly have been better if the test had been given at the end of the school year. Apparently about 85% of the juniors taking geometry had lost many of the algebra skills covered in Algebra I. Another concern was that almost 40% of the juniors who were taking "senior" math (presumably such students had at least Algebra I, Geometry, and Algebra II in earlier grades) placed at the remedial level. However, 18 of the 38 students tested as "senior" math students were at the same school and 10 of these students placed at level 4. Thus, performance at a single school dominates these data.

The section of the examination determining readiness for calculus (MPL 1) was not given. It is interesting that the high school teachers who examined the placement test given to juniors felt that most of their Algebra II students (midway through the course) should have been able to score 50% or better, thus placing at least in MPL 3.

Because the testing at Westland High School was done in two years and included two successive junior classes, the teachers and counselors there were able to observe some unanticipated effects on students. Between the spring registration of the first testing year and the beginning of classes the following September, many students changed their minds about taking an additional mathematics course (see Table C). The guidance staff at Westland feels that in most cases the decisions not to take more mathematics came from more realistic thinking about careers. Early mathematics placement testing showed many students what their proposed careers involved and helped them decide how seriously they were interested in those careers.

TABLE C

Enrollment of Seniors in Mathematics Courses—Westland High School

	1977–1978 Actual	1978–1979 Anticipated (Juniors indicating their senior course selection)	1978–1979 Actual
General Math	13	8	7
Algebra, Part I	0	7	5
Algebra, Part II	3	4	4
Consumer Math	14	13	10
Informal Geometry	5	10	6
Algebra 2	6	25	25
Senior Review Math	22	42	27
Geometry	2	0	2
Advanced Math	15	31	25
Advanced Math/Physics	29	48	39
Calculus	1	2	2
	110	190	152
Class Size	469		484

The Department of Mathematics at Ohio State has extended the Early Mathematics Placement Testing project to include thirty Ohio high schools in 1979–1980. In addition, project personnel are analyzing the university mathematics placement and course performance of the first Westland pilot group. In the autumn quarter, 1979, fifty of the students who were tested as Westland juniors enrolled at the university. Table D compares the mathematics placement levels of the 1979 freshmen from Westland with the students who came to Ohio State from Westland during the five years before 1979.

TABLE D

Mathematics Placement OSU Freshmen from Westland High School

	Five Years Before 1979		Autumn 1979	
	N	%	*N*	%
MPL 1	6	2.5	7	14.0
MPL 2	41	17.2	13	26.0
MPL 3	64	26.9	10	20.0
MPL 4	47	19.7	3	6.0
MPL 5	80	33.6	17	34.0
N	238		50	
Mean High School Rank	73.8 percentile		76.3 percentile	
Mean Math ACT	18.1		19.5	
Mean Math Placement	3.65		3.20	

The project director for the Early Mathematics Placement Testing project at The Ohio State University is Professor Bert Waits. Further information may be obtained by writing to him at the Department of Mathematics, The Ohio State University, 231 W. 18th Avenue, Columbus, Ohio 43210.

AN ELEMENTARY COURSE IN MATHEMATICAL SYMMETRY

BRUCE I. ROSE

Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556

ROBERT D. STAFFORD

Department of Mathematics, University of Houston, Houston, TX 77004

Introduction. Although a course in the mathematics of symmetry is appealing to both students and instructors, the literature, with the exception of the excellent article of Schattschneider **[10]**, does not lend itself well to the design of a course. We outline a course in the mathematics of symmetry that presupposes little or no previous contact with mathematics. Our discussion should enable a mathematics instructor without experience in this subject to teach such a course.

Even the mathematically sophisticated find identification of plane symmetry groups a time-consuming and tedious process, but our Algorithm III provides a pathway to the identification of these groups that avoids the inherent difficulties well enough so that anyone can identify planar groups. This algorithmic approach gives the student an immediate familiarity with the symmetry groups. The investigation of the mathematics of symmetry can then proceed in several directions. The study of the geometric constraints imposed on symmetric designs and the resultant structural richness of their symmetry groups can be pursued with as much rigor as is appropriate. The identification algorithms may be used "in reverse" to simplify the production of patterns with specific symmetry groups; we have found this to be a rewarding and productive topic. Unlike many enrichment courses, this course shows the student how an understanding of mathematics modifies one's view of the world. Many students are delighted by this outcome.

We use the term "frieze groups" for symmetry groups of designs on an infinite strip, and "wallpaper groups" for the symmetry groups of designs in the plane. See **[2]**, **[3]**, **[4]**, **[10]**, and **[11]**.

Course Outline. The course has four distinct parts; each part establishes a background for the next. We begin by focusing on the two interests of the course—mathematical structures and the symmetry of graphic designs. To motivate interest in mathematical structures one can begin with a discussion of what is meant by structure in various contexts, leading to a discussion of what could be meant by the term "mathematical structure." To motivate interest in mathematical symmetry one can provide several examples of symmetric designs and, as a first homework assignment, ask the students to describe the symmetries that they see in the designs.

Time spent on a leisurely introduction to the definition of a group yields benefits later in the course. Budden **[2]** has an appropriate pace. Since the notion of a group is extremely abstract for beginning students, it is useful to present examples of unmotivated mathematical structures before focusing on more natural structures. This seems to be mentally liberating for the students.

Again, in presenting the cyclic and dihedral groups one must fight the temptation to move too quickly. We found it valuable to emphasize parity arguments in constructing multiplication tables. (That is, rotations preserve orientation [proper or direct isometries]; reflections reverse orientation [improper or opposite isometries]. The product of two improper isometries is proper;

therefore, the product of two reflections must be a rotation, and so on.) At this point we chose to "prove" that any isometry taking a bounded figure to itself has a fixed point, if an isometry has two fixed points then it is a reflection, and if a symmetry group has exactly n rotations and a reflection then it has exactly n reflections.

Our students found the notion of group isomorphism to be quite difficult. After they learned to identify the symmetry groups of bounded figures and adjusted to the abstractness of the notion of isomorphism, we proceeded to examine the subgroup relationships between the cyclic and dihedral groups. Interest in this topic was stimulated by asking questions such as: Given two ornaments, what is the symmetry group of the figure obtained by imposing one design upon the other? Subgroup relationships can also be illustrated by asking students to modify a design by adding lines to obtain designs with different symmetry groups.

With the frieze groups, the students can begin to look for decorative uses of mathematical symmetry. We sought proficiency in the use of Algorithm II before examining theoretical aspects of these groups. Once this background was established, we elaborately constructed a table showing how different isometries multiply together using parity arguments. We could then examine the relationship between the minimal translation distance taking the pattern to itself and the other isometries in the symmetry group of the pattern. For example, it is useful to know that the distance between pairs of vertical reflections (and pairs of half-turns) is one-half the distance of the minimum translation that the pattern permits. Also, if a pattern contains both half-turns and vertical reflections, then the half-turns lie either on the mirror lines or halfway between adjacent mirror lines. Again, the work done here enhances the student's appreciation of the wallpaper groups.

Before presenting the algorithm for identifying the symmetry groups of the plane, a slight change in emphasis is necessary. For purposes of identification the students must focus on mirror lines rather than reflections, and glide lines rather than glide reflections. (A glide line is the mirror axis of a glide reflection; the translation component of a glide reflection is along the glide line.) The identification algorithm can be broken into four parts. For each part, students should be presented with many examples for which they have to identify the symmetry group as one of those studied or "other." The following order is suggested:

(a) No mirror lines are present—pg, ppg, p1, p2, p3, p4, p6.

(b) There exist mirror lines that are not parallel and there are *no* centers of rotation of order 6 or 3—pmm, cmm, p4m, p4g.

(c) There exist mirror lines that are not parallel and there are centers of rotation of order 6 or 3—p6m, p31m, p3m1.

(d) All mirror lines are parallel—pm, cm, pmg.

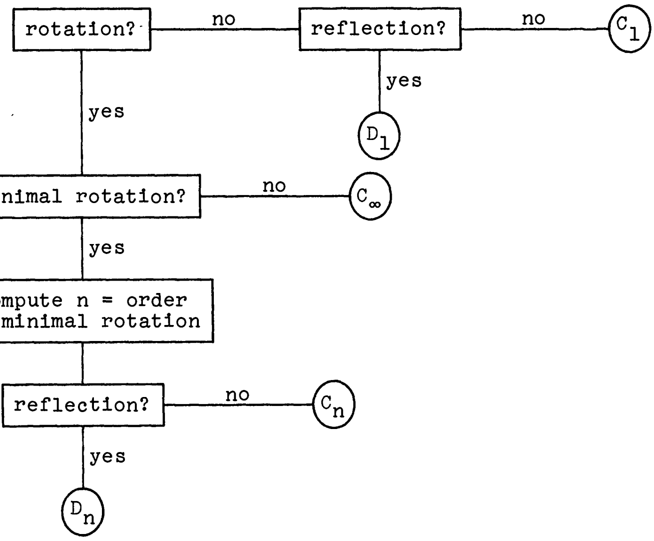
The next section examines some of the difficulties likely to be encountered in presenting this material.

At the end of the course one could examine relatively deep mathematical topics, perhaps even subgroup relationships of the wallpaper groups (cf. Coxeter and Moser [5]). However, students will probably find it more valuable to experiment with constructing patterns with different symmetry groups and perfecting their ability to identify the symmetry groups of patterns.

A detailed syllabus is available from the authors.

Algorithms. Algorithm I identifies the symmetry group of a bounded figure; Algorithm II identifies the symmetry group of a design on an infinite strip; and Algorithm III identifies the symmetry group of a design in the plane. The algorithms are designed to help mathematically unsophisticated students identify the symmetry groups. Students need to learn both a new way of thinking and a new way of seeing. Algorithm I introduces the students to an identification algorithm based on a flow chart. Algorithm II emphasizes the isometries that take a design on a strip to itself. The algorithm is organized so that the instructor can emphasize the relationships between elements of the symmetry group. Since these relationships also exist in the wallpaper groups, but in a more complicated form, a thorough examination of the frieze groups helps the

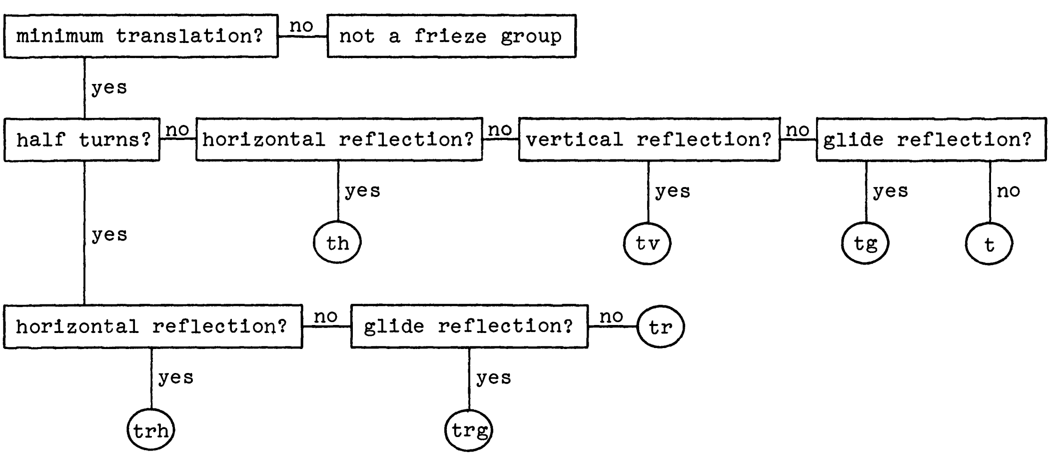
Algorithm I. Symmetry Groups of Bounded Figures



student to understand the wallpaper groups. The algorithm for the identification of the wallpaper groups mainly hinges on the student's ability to identify mirror lines—a comparatively easy ability to develop. The algorithm also employs and thus reinforces the student's recently acquired ability to identify the symmetry group of a bounded figure.

We emphasize two further virtues of Algorithm III. The first is that it directly teaches a new way of seeing designs in a plane. For example, if a figure has mirror lines all of which are parallel, then any glide lines must be either perpendicular to the mirror lines or parallel to the mirror lines (and halfway between adjacent mirror lines). The second virtue is that the algorithm can be worked backwards. This is most striking in the case of the groups pmm, cmm, p4m, and

Algorithm II. Identification Algorithm for the Frieze Groups

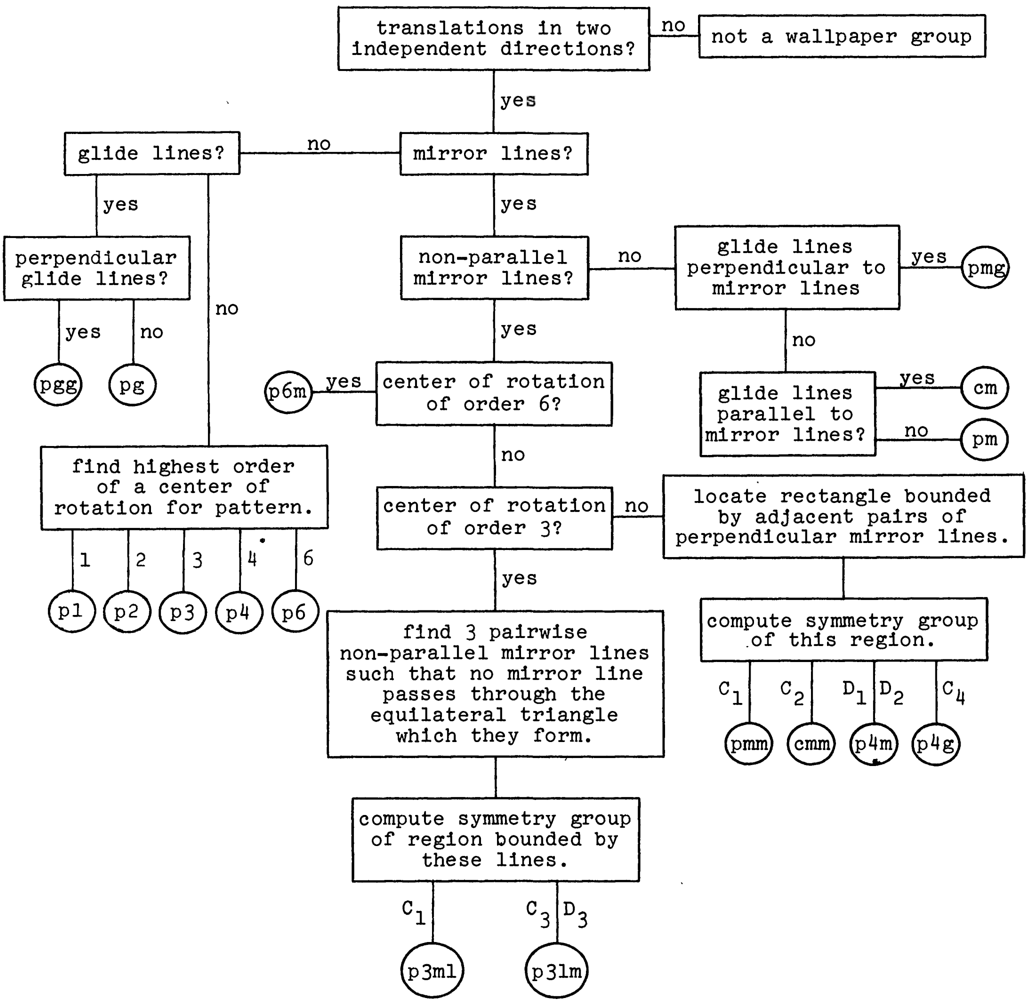


Key: t — translation
r — rotation/half turn
h — horizontal reflection
v — vertical reflection
g — glide reflection

p4g. To draw a design having one of those symmetry groups, construct a rectangular grid, draw an ornament with indicated symmetry group, and reflect the design using the grid lines as mirror lines. Thus, the algorithm serves as an introduction to the construction of intricate designs. Furthermore, there is a strong interaction between theoretical information about the relationship of elements of the symmetry group and drawing a design. These aspects of the material fascinate students who are interested in graphic design.

Algorithm I uses the standard notation for the cyclic and dihedral groups, but there is no standard notation for the frieze groups. In Algorithm II we have adopted a notation related to

Algorithm III. Identification Algorithm for the Wallpaper Groups



the algorithm for identifying a particular group. The notation used to identify the wallpaper groups (Algorithm III) is an adaptation of the symbolism used by crystallographers. See [**10**, p. 443].

Algorithms I and II are self-explanatory. We offer a few comments to facilitate the use of Algorithm III. The most important is that mirror lines (and glide lines) may be viewed as occurring in pairs. Given a mirror line, there must be a mirror line parallel to it at a distance equal to half the minimum distance one can translate the pattern in a direction perpendicular to the original mirror line. Inability to locate adjacent mirror lines may result in incorrect identification of the groups pmm, cmm, p4m, and p4g. In identifying pmm, cmm, p4m, p4g,

p31m, and p3m1, one must first identify the symmetry group of a design contained in a rectangle or triangle. One must compute the symmetry group of the part of the planar design which lies in the given figure, *including* the part of the design lying on the boundary.

In the algorithm, all glide reflections (glide lines) are intended to be nontrivial glide reflections. A glide reflection is nontrivial if its translation component and reflection component are not elements of the symmetry group.

Students will at times incorrectly identify the symmetry group of a design because certain isometries are so striking that the students ignore the sequence given in the algorithm. For example, one must locate mirror lines before searching for glide lines. Designs with symmetry group p4m may be confusing because such designs contain three sets of mirror lines. Confusion is avoided by adhering to the algorithm and examining the order of centers of rotation immediately after finding a rectangular grid of mirror lines. After the students have become adept at identification, they can investigate the individual groups in more detail. For example, they will learn that cmm contains a nontrivial glide reflection, although this fact is irrelevant to the identification algorithm.

Materials. It is easier than one might imagine to prepare materials to teach a course like this. Some combination of overhead, opaque, and slide projectors is necessary. Class interest is heightened by using natural examples, especially materials with color. Pieces of cloth and samples of wallpaper are easily obtainable. Very beautiful examples of all symmetry groups may be found in [7]. If possible, have color slides made of an assortment of the plates. They will project well onto a blackboard. One can then write on the pattern with chalk.

The difficulty of creating examples illustrating the wallpaper groups can be avoided entirely by using the paperback [6]. One can copy the designs onto transparencies or project them onto a blackboard using an opaque projector. At times it is useful to trace the projected design on the blackboard. For variety one can supplement designs from Dye's book with Escher designs. MacGillavry's book [8] is especially useful for this purpose. Rosen [9] contains an excellent bibliography on symmetry.

Various films can add a bit of spice to the course, for example, "Symmetry" (McGraw-Hill), "Adventures in Perception" (Netherlands Government Information Service), "Isometries" and "Dihedral Kaleidoscopes" (College Geometry Project).

Concluding Remarks. The material presented in this article can be of use to those who cannot devote a whole term to the study of symmetry. The symmetry groups provide a rich source of examples to supplement courses in group theory. The wallpaper groups, in particular, have rich structures and interrelationships and can profitably be used to investigate such topics as generators and relations or group actions and fundamental regions. The study of these groups can provide a point of entry to more advanced treatments, for example, Coxeter and Moser [5]. The study of topics in geometric algebra [1] can also be supplemented with this material.

References

1. E. Artin, Geometric Algebra, Wiley, New York, 1957.

2. F. J. Budden, The Fascination of Groups, Cambridge University Press, New York, 1972.

3. J. H. Cadwell, Topics in Recreational Mathematics, Cambridge University Press, New York, 1966.

4. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, 1961, 1969.

5. H. S. M. Coxeter and W. O. J. Moser, Generators and Relations for Discrete Groups, Springer-Verlag, New York, 1957, 1965.

6. Daniel S. Dye, A Grammar of Chinese Lattice, Harvard-Yenching Institute Monograph Series, vol. 6, Harvard University Press, Cambridge, Mass., 1937. (Reprinted as *Chinese Lattice Designs*, Dover, New York, 1974.)

7. Owen Jones, The Grammar of Ornament, Van Nostrand, Reinhold, New York, 1972. (Reproduction of same title, first published in 1856, reprinted in 1910 and 1928.)

8. Caroline H. MacGillavry, Symmetry Aspects of M. C. Escher's Periodic Drawings, Utrecht, Netherlands, 1965. (New edition is *Fantasy and Symmetry, the Periodic Drawings of M. C. Escher*, Harry Abrams, New York, 1976.)

9. Joe Rosen, Symmetry Discovered, Cambridge University Press, New York, 1975.

10. Doris Schattschneider, The plane symmetry groups: Their recognition and notation, this MONTHLY, 85 (1978) 439–450.

11. A. V. Shubnikov and V. A. Koptsik, Symmetry in Science and Art, Plenum Press, New York, 1974.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Triangle from Wythoff's Nim

S 18 [1979, 592]. *Proposed by V. E. Hoggatt, Jr., San Jose State University, and P. L. Mana, Albuquerque, New Mexico.*

Let $\{a_n\}$ be defined by $a_1=1$, $a_{n+1}=2+a_n$ if n is in $A_n=\{a_1,a_2,\ldots,a_n\}$, and $a_{n+1}=1+a_n$ if n is not in A_n. Also let $a_0=0$. For integers k and n with $0\leqslant k\leqslant n$, let $\left[\begin{matrix} n \\ k \end{matrix}\right]=a_n-a_k-a_{n-k}$. Prove that:

(a) There are an infinite number of integers m such that $\left[\begin{matrix} m \\ k \end{matrix}\right]=1$ for $0<k<m$.

(b) There are an infinite number of integers r such that

$$\left[\begin{matrix} r-s+t \\ t \end{matrix}\right]=\left[\begin{matrix} s \\ t \end{matrix}\right] \text{ for } 0\leqslant t\leqslant s\leqslant r.$$

I. *Solution by G. W. Peck, Massachusetts Institute of Technology.* Given a sequence S of 1's and 2's, let the derived sequence $D(S)$ be defined to have as kth element the length of the kth maximal block of consecutive 2's in S.

Let S be defined by $s_i=a_i-a_{i-1}$. Let $T(S)$ obey $t_1=2$, $t_i=s_i$, $i\neq 1$. Then the following facts are easily verified for the given sequence $\{a_i\}$.

1. $D(S)=S$; $T(D(S))=T(S)$
2. $\left[\begin{matrix} m \\ k \end{matrix}\right]=1$ for $0<k<m$ will hold if the terms up to the mth in $T(S)$ form a palindrome.
3. $\left[\begin{matrix} m \\ k \end{matrix}\right]=0$ for $0<k<m$ holds if the terms up to the mth in S form a palindrome. Condition

(b) of the problem is the same as $\left[\begin{matrix} m \\ k \end{matrix}\right]=0$ in this range.

4. A palindrome in $D(S)$ or $T(D(S))$ implies the existence of a longer palindrome in S or $T(S)$, respectively.
5. The first 3 elements of S form a palindrome; the first five elements of T form a palindrome.

It immediately follows that the palindromes in S or T corresponding to these last palindromes in $D^k(S)$ and $T(D^k(S))$ (D^k means the operation D iterated k times) yield infinite sequences of m values as the lengths of palindromes in S and T, respectively, and hence as values for which $\left[\begin{matrix} m \\ k \end{matrix}\right]=0, 1$, respectively, for $0<k<m$.

II. *Solution by David M. Bloom, Brooklyn College of CUNY.* It suffices to show that, for some *irrational* number x,

$$a_n=[nx] \qquad (\text{all } n). \tag{1}$$

Indeed, suppose (1) holds, and denote the fractional part of a number y by $\langle y\rangle$ ($\langle y\rangle=y-[y]$). Since x is irrational, the set $\{\langle nx\rangle : n\in Z^+\}$ has no largest or smallest member. Hence the assertion

$$\langle mx\rangle<\langle hx\rangle \quad \text{for } h=1,\ldots,m-1 \tag{2}$$

holds for infinitely many $m\in Z^+$, and the assertion

$$\langle rx\rangle>\langle kx\rangle \quad \text{for } k=1,\ldots,r-1 \tag{3}$$

holds for infinitely many $r\in Z^+$. Since (2) is equivalent to

$$\left[\begin{matrix} m \\ k \end{matrix}\right]=1 \qquad (k=1,\ldots,m-1),$$

assertion (a) of the problem follows. (3) is equivalent to $\left[\begin{matrix} r \\ k \end{matrix}\right]=0$; since the equation of part (b) is the same as $\left[\begin{matrix} r \\ s \end{matrix}\right]=\left[\begin{matrix} r \\ s-t \end{matrix}\right]$, (b) is established also.

We show that (1) holds with $x=(1+\sqrt{5})/2$. More generally, if we define $a_1=c$, $a_{n+1}=c+1+a_n$ if $n\in A_n$, $a_{n+1}=c+a_n$ if $n\notin A_n$ (where c is a positive integer), then (1) holds with

$$x=(c+\sqrt{c^2+4})/2.$$

This certainly is true for $n=1$ since $c<x<c+1$. Assume that it is true for $n=1,\ldots,k$; then $a_k=[kx]=N$ where $kx=N+\varepsilon$, $0<\varepsilon<1$. Since $x^{-1}k=(x-c)k=(N-ck)+\varepsilon$, it follows that $a_{N-ck}<k$, $a_{N-ck+1}\geqslant k$. Hence

$$\begin{aligned} k\in A_k &\Leftrightarrow a_{N-ck+1}=k \Leftrightarrow (N-ck+1)x<k+1\\ &\Leftrightarrow (x^{-1}k-\varepsilon+1)x<k+1\\ &\Leftrightarrow \varepsilon>1-x^{-1} \Leftrightarrow \varepsilon>1-(x-c)\\ &\Leftrightarrow (k+1)x>N+c+1. \end{aligned}$$

It follows that $a_{k+1}=[(k+1)x]$, completing the proof of (1) by induction.

Note: If we switch "$n\in A_n$" with "$n\notin A_n$" in the definition of a_{n+1}, then (1) holds with

$$x=\frac{(c+1)+\sqrt{(c+1)^2-4}}{2}.$$

Also solved by L. Kuipers (Switzerland) and the proposers.

Note. The proposers note that the sequence of m's of part (a) is $\{F_{2n-1}\}=1,2,5,13,\ldots$ and the sequence of r's of part (b) is $\{F_{2n}\}=1,3,8,\ldots$, where F_n is the nth Fibonacci number. Also, the ordered pairs $(a_n, n+a_n)$ are the safe-pairs of Wythoff's Nim.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA (USA) 94303, by May 31, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2863. *Proposed by D. Shelupsky, The City College, New York, N.Y.*

Let e_k, $k=0,1,2,\dots$ be a strictly increasing sequence of real numbers with $e_0=0$. Let $p_k(x)$ be the functions derived from the powers x^{e_k} by the Gram-Schmidt process on the interval $[0,1]$, that is, (a) $p_n(x)$ is a linear combination of x^{e_i}, $0\leqslant i\leqslant n$, (b) $\int_0^1 p_r(x)p_s(x)\,dx=\delta_{rs}$, and (c) the coefficient c_n of x^{e_n} is positive. Find a formula for c_n.

E 2864. *Proposed by Robert Kusner, Haverford College.*

In real affine d-space R^d let $L_n=\{A_j\}_{j=1,\dots,n}$ be a family of n hyperplanes in general position (that is, such that $\dim(\bigcap_{j\in J}A_j)=d-\operatorname{card}J$ for $0\leqslant\operatorname{card}J\leqslant d+1$). Let $C_1,\dots,C_r$ be the bounded components of the complement of $\bigcup_{1\leqslant j\leqslant n}A_j$; then each $\bar{C}_i$ is a convex polytope of dimension d with $v(\bar{C}_i)$ vertices. Let $v(L_n)=\max_{1\leqslant i\leqslant r}\{v(\bar{C}_i)\}$ and let $v(n,d)=\min_{L_n}\{v(L_n)\}$, where the minimum is taken over all families L_n of n hyperplanes in general position in R^d.

(a) Show that $v(n,2)=4$ for $n\geqslant 4$.

*(b) Decide whether $v(n,d)=2d$ for all $n\geqslant 2d\geqslant 2$.

E 2865. *Proposed by David K. Cohoon, USAF School of Aerospace Medicine, San Antonio, Texas.*

Show that for positive real a

$$\int_0^\infty\left[\ln\left(\frac{x}{x+a}\right)\right]^2 dx=\pi^2 a/3.$$

E 2866. *Proposed by Jordi Dou, University of Bàrcelona, Spain.*

Let AKL, AMN be equilateral triangles. Prove that the equilateral triangles LMX, NKY are concentric (if Y is on the properly chosen side of NK).

E 2867. *Proposed by Dennis K. Mick, Carroll College, Waukesha, Wisconsin.*

Let $h(t)>0$, $g(t)>0$ be continuous functions for $0\leqslant t<\infty$, $\int_0^\infty h(t)dt=\infty$, $\int_0^\infty g(t)dt<\infty$. Prove that there exist arbitrarily large values of r such that for any s satisfying $0\leqslant s\leqslant r$ we have $\int_{r-s}^{r+s}h(t)dt\geqslant\int_{r-s}^{r+s}g(t)dt$.

E 2868. *Proposed by Ben B. Bowen, Vallejo, California.*

Find an asymptotic estimate ($n\to\infty$, n odd) for

$$A_k(n)=\int_0^n (x-k)^{-1}\prod_{i=0}^{n}(x-i)\,dx.$$

SOLUTIONS OF ELEMENTARY PROBLEMS

A Parallelogram Generated from a Triangle

E 2802 [1979, 784]. *Proposed by M. Slater, University of Bristol, England.*

Given a triangle ABC (in the Euclidean plane), construct similar isosceles triangles ABC' and ACB' outwards on the respective bases AB and AC, and BCA'' inwards on the base BC (or

ABC'' and ACB'' inwards and BCA' outwards). Show that $AB'A''C'$ (respectively, $AB''A'C''$) is a parallelogram.

Most of the solutions used standard methods of plane geometry to argue similarity and congruence of triangles. Some more unusual solutions are presented here.

I. *Solution by Robert Young, Cape Cod Community College.* Let θ be the directed angle ACB', and $r = AC/CB'$. Let f denote the composition of the following four linear transformations: (i) magnification about C by $1/r$; (ii) rotation about C through θ; (iii) magnification about B' by r; (iv) rotation about B' through $-\theta$. Clearly f is a translation, and $f(A) = B'$ and $f(C') = A''$. Hence AB' and $C'A''$ are parallel and equal.

II. *Solution by Jordi Dou, Barcelona, Spain.* Let A_0, B_0, and C_0 be midpoints of the sides BC, CA, and AB, respectively. Note that $\overrightarrow{AA_0} = \overrightarrow{AB_0} + \overrightarrow{AC_0}$. Also, $\overrightarrow{A_0A''} = \overrightarrow{B_0B'} + \overrightarrow{C_0C'}$, since these vectors are orthogonal to, and proportional in length to, the sides of ABC. Adding the two vector equations, we obtain $\overrightarrow{AA''} = \overrightarrow{AB'} + \overrightarrow{AC'}$, which demonstrates the proposition.

L. Kuipers (Switzerland) submitted essentially the same solution, as well as four other solutions.

III. *Solution by Marlow Sholander, Case Western Reserve University.* Let points in the plane be identified with their complex coordinates. For given complex γ, let ABC be called a γ-*triangle* if $C = (1-\gamma)A + \gamma B$. All γ-triangles corresponding to a fixed γ are strictly similar. Given a quadrilateral $ABCD$ and a number γ, let points P, Q, R, and S be chosen so that ABP, CBQ, CDR, and ADS are γ-triangles, i.e.,

$$\begin{aligned} P &= (1-\gamma)A + \gamma B \\ Q &= (1-\gamma)C + \gamma B \\ R &= (1-\gamma)C + \gamma D \\ S &= (1-\gamma)A + \gamma D. \end{aligned}$$

Note that $\frac{1}{2}(P+R) = \frac{1}{2}(Q+S)$, so $PQRS$ is a quadrilateral whose diagonals bisect each other and is therefore a parallelogram.

If $\gamma = \frac{1}{2}$, we have the classical theorem concerning midpoints of sides of a quadrilateral. If γ has real part $\frac{1}{2}$ and if $D = A$, we have the proposed problem.

Sholander invites readers to establish a four-dimensional version of the result above using quaternions in place of complex numbers.

For fixed γ, the binary operation $A \# B = (1-\gamma)A + \gamma B$ generates a groupoid; it is discussed in Sholander's paper "On the Existence of the Inverse Operation in Alternation Groupoids," Bulletin of the AMS, 55 (1949) 747 (Example E). Ronald S. Tiberio refers to a paper by David Merriell, "An Application of Quasi-Groups to Geometry," this MONTHLY, 77 (1970) 44–46, where the same structure is discussed.

Clayton W. Dodge, Howard Eves, Robert Lyness (U.K.), and R. D. Nelson (U.K.) use essentially the same idea as Sholander's but apply it directly to a triangle. Eves also submitted a second solution.

Dodge, Eves, Lyness, and J. Oman note that the result follows if the triangles are merely supposed to be similar and not necessarily isosceles. Many other solvers did not require the hypothesis that the triangles be isosceles, although they did not so state explicitly.

Ken Brown and Rodney T. Hood note that the parallelogram may be degenerate.

Also solved by Dieter Bode (Germany), Béla Brindza (Hungary), Ragnar Dybvik (Norway), Henry E. Fettis, Barbara J. Gaitley, Samuel L. Greitzer, O. P. Lossers (Netherlands), Hubert J. Ludwig, Russell Lyons, Nicholas A. Martin, H. Stephen Morse, Ivan Paasche (Germany), Mary E. Ruddy, I. A. Sakmar (France), Jean Chan Stanek, Robert A. Sutton, Jr., Michael Vowe (Switzerland), and J. H. Webb (South Africa).

Distinct Prime Divisors of $2^k - 1$

E 2805 [1979, 864]. *Proposed by Wells Johnson, Bowdoin College, Maine.*

Let the integer $r \geqslant 0$ be given. Show that each of the numbers $(2^{2^r})^n - 1$ has at least $2r+1$

distinct prime factors if $n > 2^r$, with the lone exception $r=1$, $n=3$, when $4^3-1=3^2\cdot 7$.

Solution by L. E. Mattics, University of South Alabama. Let $a>1$, $b>1$, $n>1$; $(a,b,n)\neq(3,2,2)$. Suppose further that for $1\leqslant i\leqslant r$, $n\neq a^i$. Then $N=(b^{a^r})^n-1$ is divisible by at least $2r+1$ distinct primes, except in the cases $(a,b,n,r)=(2,2,3,1),(2,2,3,2),(2,2,3,3)$. [The assertion of the proposal is the case $(a,b,n)=(2,2,n)$, $(n,r)\neq(3,1)$.] Using G. D. Birkhoff and H. S. Vandiver, *On the integral divisors of* a^n-b^n, Annals of Math., (2) 5 (1904) 173–180, we note that each of the k numbers

$$b^a-1,\ldots,b^{a^r}-1,(b)^n-1,(b^a)^n-1,\ldots,(b^{a^r})^n-1$$

possesses a prime factor not dividing any of the others except possibly when $(a,b,n)=(3,2,2),(2,2,3)$. Also, each of the k numbers is a factor of N. But when $(a,b,n)=(2,2,3)$, $r=4$, $97\cdot 673$ divides $2^{48}-1$, but divides none of the other eight numbers [$k=9$]. In the context, the cases $(a,b,n)=(2,2,3)$, $r>4$ are also settled by this remark.

The case $(a,b,n)=(3,2,2)$ may be similar, but the value of r ($r>0$) required to generate $2r+1$ distinct prime factors is undecided.

Also solved by H. L. Abbott, A. Adelberg, W. Boucher (Canada), R. Breusch, D. M. Broline, P. Filakovsky (Hungary), L. L. Foster, E. R. Gentile (Argentina), L. Jones, J. Leech, M. R. Modak (India), D. Orr, H. Schmidt, Jr., G. Shulman, G. J. Simmons, R. Teitler (U.K.), E. Trost (Switzerland), and the proposer.

Abbott referred to W. Sierpiński, *Theory of Numbers*, pp. 335, 346.

Inequality $(n+a)^k\leqslant rn^k$

E 2807 [1979, 865]. *Proposed by Solomon W. Golomb, University of Southern California.*

Let a and r be fixed positive constants with $r>1$. For each positive integer k there is a smallest positive integer $n=n(k)$ which satisfies $(n+a)^k\leqslant rn^k$. Show that $\lim n(k)/k$ as $k\to\infty$ exists and evaluate this limit.

Solution by Nick Franceschine III, Sebastopol, California. Two related results are: (i) if the inequality of the proposal is replaced by $(bn+a)^k\leqslant r(bn+c)^k$, $r>1$, $b>0$, $a>c\geqslant 0$, then $n(k)/k\to(a-c)/(b\ln r)$ as $k\to\infty$; (ii) if the inequality of the proposal is replaced by $(n^2+bn+a)^k\leqslant r(n^2+bn+c)^k$, $r>1$, $b>0$, $a>c\geqslant 0$, then $n(k)^2/k\to(a-c)/\ln r$ as $k\to\infty$. The proposal comes from replacing b, c in (i) by 1,0. We prove (i). From $(bn+a)^k\leqslant r(bn+c)^k$, it follows that $n\geqslant(a-cr^{1/k})/[b(r^{1/k}-1)]$. Thus $|n(k)/k-(a-cr^{1/k})/[b(r^{1/k}-1)]|<1/k$. The assertion follows from application of L'Hôpital's rule: $k(r^{1/k}-1)\to\ln r$, as $k\to\infty$.

Also solved by H. L. Abbott (Canada), K. Béla (student, Hungary), T. Q. Binh (student, Hungary), W. Boucher (Canada), D. M. Bloom, R. Breusch, Béla Brindza (Hungary), K. Brown (two solutions), F. S. Cater, Chico Problem Group, E. Dixon, M. J. Dixon, U. Faigle (Germany), P. Filakovsky (student, Hungary), L. L. Foster, E. R. Gentile (Argentina), S. J. Goodenough & T. M. Mills (Australia), E. Grosswald, H. Kappus (Switzerland), M. F. Kruelle (student), L. Kuipers (Switzerland), J. Leech, R. A. Leslie, J. Levy, O. P. Lossers, D. K. Mick, L. T. Phat, E. Posti (Finland), P. Schumer, D. L. Shell, E. Simko (student, Czechoslovakia), A. Smuckler (Israel), J. Tripp, University of South Alabama Problem Group, and the proposer.

ADVANCED PROBLEMS

Solutions of Advanced Problems should be sent to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA). Solutions of Advanced Problems in this issue should be typed (in duplicate, with double spacing) and should be mailed before May 31, 1981. If acknowledgment is desired, include a self-addressed card.

6326*. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

A topological space S is *quasi-connected* provided that any covering of S by disjoint nonempty open sets is finite. Equivalently, any continuous map from S into a discrete space has

finite range. It is not hard to show that, if S is quasi-connected and T is either connected or compact, then $S\times T$ is quasi-connected.

Is the product of two quasi-connected spaces always quasi-connected? What about infinitely many?

6327. *Proposed by Tamás F. Móri and S. Szántó, Lórand Eötvös University, Budapest, Hungary.*

Are the random variables X and Y independent if they are conditionally uncorrelated in the following sense:

$$E(XY|\, a\leqslant X<b, c\leqslant Y<d)=E(X|\, a\leqslant X<b, c\leqslant Y<d)E(Y|\, a\leqslant X<b, c\leqslant Y<d)$$

for arbitrary real numbers a, b, c, d having the property

$$P(a\leqslant X<b, c\leqslant Y<d)>0?$$

*The same question, if the intervals $[a,b)$, $[c,d)$ are replaced by $(-\infty,b)$, $(-\infty,d)$.

6328*. *Proposed by N. P. Erugin, Minsk, USSR.*

The equation $\dot{x}=P(x,y)$, $\dot{y}=Q(x,y)$ can be solved in closed form to within quadratures if P,Q are homogeneous polynomials of the same degree, or if P,Q are both polynomials of first degree. If P,Q are both of degree at most 2, will there be closed form solutions in other cases? If so, in what other cases?

6329. *Proposed by Ronald J. Evans, University of California at San Diego.*

Let $p>3$ be prime. When $p\equiv 1 \pmod 6$, write $p=c^2+3d^2$. Let $\chi(m)$ denote the Legendre symbol (m/p). Define $S=\sum\prod_{i=1}^{4}\chi(x_1x_2x_3x_4+x_i)$, where the sum is extended over all x_1,x_2,x_3,x_4 (mod$p$). Show that $S=1+2p^2-16pc^2+16c^4$, if $p\equiv 1 \pmod 6$, and $S=1$, if $p\equiv -1 \pmod 6$. Also, evaluate $\sum\prod_{i=1}^{3}\chi(x_1x_2x_3+x_i)$, where the sum is over all x_1,x_2,x_3 (mod$p$).

SOLUTIONS OF ADVANCED PROBLEMS

Collineations of Projective Spaces

6236 [1978, 770] and 6267 [1979, 398]. *Proposed by Antal E. Fekete, Memorial University of Newfoundland.*

6236: We say that two endomorphisms of the complex vector space C^n are of the same type if there is a bijection between their respective sets of eigenvalues which maps the Jordan normal form of one endomorphism into that of the other. Find a formula determining the number of different endomorphism types of C^n. Define what is meant by endomorphism types of the real vector space R^n and determine their number.

6267: We say that two collineations of the real projective space PR^n are of the same type if their invariant configurations are projectively equivalent (i.e., there is a real projective collineation mapping one configuration into the other). Find an explicit formula determining the number of all different non-identity collineation types. For example, for $n=1$ there are 3 types: hyperbolic (two fixed points), parabolic (one fixed point) and elliptic (no fixed point). Also, define collineation types for the complex projective space PC^n and find their number.

Solution to both problems by R. K. Oliver, Pittsburgh, PA. Let c_n (respectively, r_n) denote the number of types of collineations (including the identity) of $P(C^n)=PC^{n-1}$ (respectively, $P(R^n)=PR^{n-1}$), $n\geqslant 1$, and put $c_0=r_0=1$. Let p_n denote the number of partitions of $n, n\geqslant 1$.

Then $\Sigma_{n\geq 0} c_n x^n = \Pi_{m\geq 1}(1-x^m)^{-p_m}$ and $\Sigma_{n\geq 0} r_n x^n = (\Sigma_{n\geq 0} c_n x^n)(\Sigma_{n\geq 0} c_n x^{2n})$. The first formula, which is due to Cayley [1] (cf. also [2]), follows easily from the fact that the number of types of Jordan normal forms of elements of $GL(n,C)$ equals the number of what might be called second-order partitions of n (formed by first partitioning n and then partitioning the resulting parts) and the fact that the number of k-selections of the set $\{1,\ldots,p_m\}$ equals the number of p_m-compositions of the integer k, $k\geq 1$. (Here k-selection of $\{1,\ldots,p_m\}$ means a nondecreasing sequence of length k with values in $\{1,\ldots,p_m\}$, p_m-composition of k means a sequence of nonnegative integers with length p_m and sum k, and k is to be interpreted as the number of parts of size m in the partition of n.) And the second formula follows immediately from the observation that in the Jordan normal form of an element of $GL(n,R)$ the blocks corresponding to the complex eigenvalues occur in conjugate pairs. For $1\leq n\leq 11$ the numbers c_n and r_n are, respectively, 1,3,6,14,27,58,111,223,424,817,1527, and 1,4,7,20,36,87,162,355,666,1367,2557.

References

1. A. Cayley, Recherches sur les matrices dont les termes sont des fonctions linéaires d'une seule indéterminée, J. Reine Angew. Math., 50 (1855) 313–317 = Collected Mathematical Papers, vol. 2, pp. 216–220.

2. T. J. I'A. Bromwich, Quadratic forms and their classification by means of invariant-factors, Cambridge, 1906, p. 60.

Concentrated Sets of Reals

6261 [1979, 226]. *Proposed by Hugh Noland, California State University, Chico.*

Let S be an uncountable set of real numbers and let A be a countable subset of S. Must there exist an open set U, containing A, such that $S-U$ is uncountable?

Solution by Curtis D. Herink, Allegheny College. The answer to this problem is independent of the usual axioms of set theory. This can most easily be shown by finding two axioms, each of which is already known to be consistent with the usual axioms of set theory, such that the first implies the answer is no and the second implies the answer is yes. The appropriate axioms are the continuum hypothesis, and Martin's axiom plus the negation of the continuum hypothesis.

A. The continuum hypothesis implies that there is an uncountable set S of real numbers and a countable subset A of S such that every open set containing A contains all but countably many points of S. When this happens, we say that S is *concentrated* on A.

It is well known that the continuum hypothesis implies the existence of a Luzin set, i.e., an uncountable set of reals whose intersection with any nowhere dense set is countable. (See N. Luzin, *Sur un problème de M. Baire*, C. R. Acad. Sci. Paris, 158 (1914) 1258–1261, or Kenneth Kunen, "Combinatorics," in Jon Barwise, ed., *Handbook of Mathematical Logic*, Amsterdam, 1977, pp. 371–401.) We fix a Luzin set S and find a countable subset A of S such that S is concentrated on A. In fact, any countable subset of S whose closure contains S will do. It is easy to find such sets because the reals have a countable basis. If U is any open set containing A, it is easy to check that S/U is nowhere dense and therefore countable.

B. Martin's axiom plus the negation of the continuum hypothesis implies that no uncountable subset of the reals can be concentrated on any countable set. For a definition of terms used and a statement of Martin's axiom, see J. R. Shoenfield, *Martin's Axiom*, this MONTHLY, 82 (1975) 610–617. (Some of these terms are not in general use. In particular, subnets are usually referred to as filters.)

Let S be uncountable and $A=\{a_n : n\in\omega\}$ be countable. Choose distinct points c_α $(\alpha<\omega_1)$ belonging to $S\backslash A$. Further for each $a_n\in A$, choose a neighborhood basis $\mathbf{B}_n=\{B_{n,m} : m\in\omega\}$ for a_n such that

(1) $m'>m \Rightarrow B_{n,m'}\subset B_{n,m'}$, and
(2) there are uncountably many c_α not belonging to $\bigcup_{i<n} B_{i,0}$.

Now let P be the set of all ordered pairs (f,X) such that, for some integer k, f is a function with domain $\{0,1,\ldots,k\}$ and $f(n)\in\mathbf{B}_n$ for all $n\leqslant k$ and X is a finite collection of the c_α disjoint from $\bigcup_{n\leqslant k}f(n)$. Partially order P by $(f,X)\geqslant(f_1,X_1)$ iff f extends f_1 and $X\supset X_1$. Since there are only countably many functions f and any two elements (f,X) and (f,Y) of P with the same first element are compatible, $(P,\leqslant)$ has the countable chain condition and we may apply Martin's axiom to it. Define the following subsets of P:

$$D_n=\{(f,X)\in P: n\in\operatorname{dom} f\}\qquad (n<\omega)$$

$$E_\alpha=\{(f,X)\in P:\text{for some }\beta\geqslant\alpha,\ a_\beta\in X\}\qquad(\alpha<\omega_1).$$

It is easy to see that each D_n and E_α is dense in $(P,\leqslant)$. By Martin's axiom, let G be a subnet of P with nonempty intersection with each D_n and E_α. Since any two members of a subnet are compatible, it is easy to verify that:

(1) $\cup\{f(n):\text{ for some }X,\ (f,X)\in G\text{ and }n\in\operatorname{dom} f\}=U$ is an open set containing A,

(2) $\{c_\alpha:\text{ for some }(f,X)\in G,\ c_\alpha\in X\}=C$ is an uncountable subset of S, and

(3) C is disjoint from U. Hence S is not concentrated on A.

Also solved by Charles L. Belna, F. S. Cater, James W. Fickett, John C. Morgan II, Ivan Netuka (Czechoslovakia), Stephen Noltie, Nicholas Passell, Paul Perlmutter, Roderick A. Price, Richard B. Tucker, Erik K. van Douwen, Stanley Wagon, and the proposer.

Solvers provided the following additional relevant references.

1. A. S. Besicovitch, Concentrated and rarified sets of points, Acta Math., 62 (1934) 289–300.
2. R. Laver, On the consistency of Borel's conjecture, Acta Math., 137 (1976) 151–169.
3. C. A. Rogers, Hausdorff Measures, Cambridge Univ. Press, 1970, esp. Theorem 38, p. 74.
4. F. Rothberger, Sur les familles indénombrables de suites de nombres naturels et les problèmes concernant la propriété C., Proc. Cambridge Phil. Soc., 37 (1941) 109–126.

Fixed Points of Trees

6262 [1979, 226]. *Proposed by A. Blass and F. Harary, University of Michigan, and W. T. Trotter, Jr., University of South Carolina.*

What is the probability that a tree selected at random has a fixed point? More specifically, let t_n be the number of (nonisomorphic) trees with n points and let f_n be the number of such trees T with at least one point fixed under all automorphisms of T. Calculate $\lim_{n\to\infty} f_n/t_n$.

Solution by Lajos Takács, Case Western Reserve University. The answer is 1. In what follows, T denotes a tree with n vertices, σ, an automorphism of T, and r_n, the number of nonisomorphic rooted trees with n vertices.

We observe that if T consists of two isomorphic rooted trees, T_a and T_b, whose roots, a and b, are joined by an edge, and if $\sigma(T_a)=T_b$, then σ has no fixed points. This can happen only if n is even, and the number of such trees T is $r_{n/2}$. Conversely, if σ has no fixed points, then T is of the above form. For, by a theorem of R. Otter (1948), if σ has no fixed points, then T has an edge (a,b) for which $\sigma(a)=b$ and $\sigma(b)=a$. By removing the edge (a,b), the tree T splits into two trees T_a and T_b in such a way that a is in T_a, b is in T_b, and $\sigma(T_a)=T_b$. If we assign a and b as the roots of T_a and T_b, respectively, then T_a and T_b become isomorphic rooted trees.

Accordingly, $f_n=t_n$ if n is odd, and $f_n=t_n-r_{n/2}$ if n is even. We have $\lim_{n\to\infty}f_n/t_n=1$ if and only if $\lim_{m\to\infty}r_m/t_{2m}=0$. R. Otter determined the asymptotic behavior of r_n and t_n as $n\to\infty$, and his results imply that $\lim r_m/t_{2m}=0$. However, this can also be proved in an elementary way. By a result of A. Cayley (1889) the number of distinct trees with n vertices is n^{n-2} and thus $(n-1)!\,r_n\geqslant n^{n-2}$, or $r_n\geqslant n^{n-1}/n!\geqslant e^{n-1}n^{-3/2}$ for $n\geqslant 1$. The last inequality is a consequence of the following one:

$$\log 1+\cdots+\log n\leqslant\int_1^n\log x\,dx+\tfrac{1}{2}\log n=n(\log n-1)+\tfrac{1}{2}\log n+1.$$

Since, obviously, $r_m^2\leqslant r_{2m}\leqslant 2mt_{2m}$, we have

$$0 \leqslant \frac{r_m}{t_{2m}} \leqslant \frac{2mr_m}{r_{2m}} \leqslant \frac{2m}{r_m} \leqslant \frac{2m^{5/2}}{e^{m-1}} \to 0.$$

References. A. Cayley, A theorem on trees, Quart. J. of Pure and Appl. Math., 23 (1889) 376–378. The Collected Mathematical Papers of Arthur Cayley, vol. 13, Cambridge University Press, 1897, pp. 26–28.

R. Otter, The number of trees, Ann. of Math., 49 (1948) 583–599.

Also solved by P. J. Federico and the proposers.

Inequality Involving the Γ Function

6269 [1979, 399]. *Proposed by Robert E. Shafer, Berkeley, California.*

Let $F(u)=u^{-u}\Gamma(u+\frac{1}{2})$ and

$$G(x,s,t)=\frac{1}{(x-s+\frac{1}{2})(x-t+\frac{1}{2})}-\frac{1}{(x+s+\frac{1}{2})(x+t+\frac{1}{2})}.$$

Prove that, for $0 \leqslant s < t \leqslant x$,

$$e^{(s-t)G(x,s,t)/24} < \frac{F(x-t+\frac{1}{2})F(x+t+\frac{1}{2})}{F(x-s+\frac{1}{2})F(x+s+\frac{1}{2})} < 1.$$

Solution by L. Kuipers, Mollens, Switzerland. First we show:

$$\frac{(x-t+\frac{1}{2})^{-x+t-\frac{1}{2}}(x+t+\frac{1}{2})^{-x-t-\frac{1}{2}}\Gamma(x-t+1)\Gamma(x+t+1)}{(x-s+\frac{1}{2})^{-x+s-\frac{1}{2}}(x+s+\frac{1}{2})^{-x-s-\frac{1}{2}}\Gamma(x-s+1)\Gamma(x+s+1)} < 1,$$

or

$$\log\Gamma(x-t+1)+\log\Gamma(x+t+1)-(x-t+\tfrac{1}{2})\log(x-t+\tfrac{1}{2})-(x+t+\tfrac{1}{2})\log(x+t+\tfrac{1}{2})$$
$$< \log\Gamma(x-s+1)+\log\Gamma(x+s+1)-(x-s+\tfrac{1}{2})\log(x-s+\tfrac{1}{2})-(x+s+\tfrac{1}{2})\log(x+s+\tfrac{1}{2}).$$

Denote the left-hand side of this inequality by $g(t)$. We want to show:

$$g(t)-g(s)<0 \qquad (0 \leqslant s < t \leqslant x).$$

If $t=x$, we have $g(t)-g(s)=0$.

Now differentiate with respect to t. If the derivative $(t \geqslant s)$ is negative, we are done. We obtain:

$$-\frac{\Gamma'(x-t+1)}{\Gamma(x-t+1)}+\frac{\Gamma'(x+t+1)}{\Gamma(x+t+1)}+\log(x-t+\tfrac{1}{2})-\log(x+t+\tfrac{1}{2})$$
$$=-\psi(x-t+1)+\psi(x+t+1)+\log(x-t+\tfrac{1}{2})-\log(x+t+\tfrac{1}{2}),$$
$$=h(t),$$

say. Recall $0 \leqslant t \leqslant x$. Obviously $h(0)=0$.

Furthermore:

$$h'(t)=\psi'(x-t+1)+\psi'(x+t+1)-\frac{1}{x-t+\frac{1}{2}}-\frac{1}{x+t+\frac{1}{2}}$$
$$=\sum_{n=0}^{\infty}\frac{1}{(x-t+1+n)^2}+\sum_{n=0}^{\infty}\frac{1}{(x+t+1+n)^2}-\frac{1}{x-t+\frac{1}{2}}-\frac{1}{x+t+\frac{1}{2}}>0,$$

since

$$\frac{1}{(a+1)^2}+\frac{1}{(a+2)^2}+\cdots<\frac{1}{a+\frac{1}{2}} \qquad (a<0).$$

So $h(t)<0$ $(t>0)$ and $g(t)-g(s)<0$. We have to show:

$$(s-t)G(x,s,t)/24<\log F\left(x-t+\tfrac{1}{2}\right)+\log F\left(x+t+\tfrac{1}{2}\right)-\log F\left(x-s+\tfrac{1}{2}\right)-\log F\left(x+s+\tfrac{1}{2}\right).$$

The left-hand side of this inequality can be written in the form:

$$\frac{1}{24}\left\{\frac{1}{x-s+\frac{1}{2}}+\frac{1}{x+s+\frac{1}{2}}-\frac{1}{x-t+\frac{1}{2}}-\frac{1}{x+t+\frac{1}{2}}\right\}.$$

Write:

$$k(t)=\log F(x-t+1)+\log F(x+t+1)+\frac{1}{24\left(x-t+\frac{1}{2}\right)}+\frac{1}{24\left(x+t+\frac{1}{2}\right)}.$$

Then we want to show:

$$k(t)-k(s)>0.$$

The derivative of $k(t)$, or of

$$-\left(x-t+\tfrac{1}{2}\right)\log\left(x-t+\tfrac{1}{2}\right)-\left(x+t+\tfrac{1}{2}\right)\log\left(x+t+\tfrac{1}{2}\right)+\log\Gamma(x-t+1)$$
$$+\log\Gamma(x+t+1)+\frac{1}{24\left(x-t+\frac{1}{2}\right)}+\frac{1}{24\left(x+t+\frac{1}{2}\right)},$$

is equal to:

$$\log\left(x-t+\tfrac{1}{2}\right)-\log\left(x+t+\tfrac{1}{2}\right)-\psi(x-t+1)+\psi(x+t+1)+\frac{1}{24\left(x-t+\frac{1}{2}\right)^2}-\frac{1}{24\left(x+t+\frac{1}{2}\right)^2}=l(t),$$

say. For $t=0$ we have $l(t)=0$.

Now

$$l'(t)=\psi'(x-t+1)+\psi'(x+t+1)-\frac{1}{x-t+\frac{1}{2}}-\frac{1}{x+t+\frac{1}{2}}+\frac{1}{12\left(x-t+\frac{1}{2}\right)^3}+\frac{1}{12\left(x+t+\frac{1}{2}\right)^3}$$
$$=\sum_{n=0}^{\infty}\frac{1}{(x-t+1+n)^2}+\sum_{n=0}^{\infty}\frac{1}{(x+t+1+n)^2}-\frac{1}{x-t+\frac{1}{2}}+\frac{1}{x+t+\frac{1}{2}}$$
$$+\frac{1}{12\left(x-t+\frac{1}{2}\right)^3}+\frac{1}{12\left(x+t+\frac{1}{2}\right)^3}>0,$$

since

$$\frac{1}{(a+1)^2}+\frac{1}{(a+2)^2}+\cdots>\frac{1}{a+\frac{1}{2}}-\frac{1}{12\left(a+\frac{1}{2}\right)^2}.$$

Hence $l'(t)>0$ implies $k'(t)>0$, which implies $k(t)>k(s)$.

Also solved by the proposer.

MISCELLANEA

50. Our knowledge is made up of the stories we can tell, stories that must be told in the language that we know. (Even mathematics is a "language" that states propositions and tells stories. It's a very elaborate form of "play" language. That's why it's such fun for those who speak it well.)

—Richard Mitchell, *Less Than Words Can Say* (Little, Brown, New York, 1979), p. 34. (Suggested by P. R. Halmos.)

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, T(13: 1, 2). The Nature of Modern Mathematics, Third Edition. Karl J. Smith. Brooks/Cole, 1980, xx + 620 pp, $16.95. [ISBN: 0-8185-0352-1] Written to create a positive attitude toward mathematics, this books asks the student to "experience math" rather than to "do math problems." Liberal use of notes, charts, anecdotes, quotes, pictures, cartoons, diagrams, etc. The text allows great flexibility in coverage and level; computers (Basic) can be integrated into the course if desired. (First Edition, TR, August-September 1975; Second Edition, TR, August-September 1976.) LCL

General, T(13-16: 1, 2), S, L.** Mathematics: Problem Solving through Recreational Mathematics. Bonnie Averbach, Orin Chein. Freeman, 1980, xvi + 400 pp, $16.50. [ISBN: 0-7167-1124-9] Recreational problems, puzzles and games for fun and motivation. Authors promote critical thinking and strategy in place of copying and manipulation. For a course in mathematics but not merely for the fun of it. Written for participants, not observers. Eight, mostly independent, chapters each built around a theme, plus an excellent potpourri ninth. Selected, but not brief, bibliography. Good index. Answers, hints, solutions to many problems. Highly recommended for students taking only one course in mathematics. JK

Precalculus, T(13: 1). Essential Precalculus. Raymond McGivney, James McKim, Benedict Pollina. Wadsworth, 1980, xi + 452 pp, $4.95. [ISBN: 0-534-00766-X] A good text for a thorough precalculus course. The careful attention to word problems should pay off for students preparing for calculus. LLK

History, P, L.** Peano: Life and Works of Giuseppe Peano. Hubert C. Kennedy. Reidel, 1980, xii + 230 pp, $14.95 (P). [ISBN: 90-277-1068-6] The first full biography of Peano, describing his life, his development of symbolic logic, his mathematics (first axiomatic definition of a vector space; a new definition of measure; numerous counterexamples) and his efforts to promote Interlingua. Appendices include lists of Peano's publications, of his students, and brief biographies of his professors. LAS

History, P, L*. A History of Computing in the Twentieth Century: A Collection of Essays. Ed: N. Metropolis, J. Howlett, Gian-Carlo Rota. Acad Pr, 1980, xix + 659 pp, $29.50. [ISBN: 0-12-491650-3] 38 papers from a June 1976 conference at Los Alamos celebrating sources of the computing revolution: "The willingness of private enterprise...the personalities of Turing and von Neumann...a few men who were immune to the stupefying demands of some presumed relevance...inventions motivated by the purest of mathematics...[and] farsighted generals and admirals... ." LAS

Foundations, S(15-18), P. A Theory of Syntactic Recognition for Natural Language. Mitchell P. Marcus. MIT Pr, 1980, 335 pp, $25. [ISBN: 0-262-13149-8] This research, based strictly on English, sets forth some principles of processing and explanations for some fundamental properties of language. MU

Combinatorics, T(18: 1), S, P, L. Eigenvalue Techniques in Design and Graph Theory. W.H. Haemers. Math. Centre Tracts, No. 121. Math Centrum, 1980, 102 pp, Dfl. 13 (P). A theorem of the author's concerning eigenvalues of partitioned Hermitian matrices is applied to obtain a number of results in graph theory, some of them new. Applications include inequalities for size of cocliques and chromatic number, determination of 4-colorable strongly regular graphs, and constructions for designs and strongly regular graphs. Bibliography, index. JS

Combinatorics, T(17: 1), S, P, L. Graphs, Codes and Designs. P.J. Cameron, J.H. van Lint. London Math. Soc. Lect. Note Ser., No. 43. Cambridge U Pr, 1980, vii + 147 pp, $18.50 (P). [ISBN: 0-521-23141-8] A collection of lectures with the aim of presenting developments in graph theory and coding theory which have a bearing on design theory. Includes an introductory chapter on designs and an extensive bibliography. CEC

Combinatorics, S(18), P. Spectra of Graphs, Theory and Application. Dragos M. Cvetković, Michael Doob, Horst Sachs. Pure and Appl. Math., V. 87. Acad Pr, 1980, 368 pp, $45. [ISBN: 0-12-195150-2] Encyclopedic treatment of spectral theory of matrices applied to graphs, a subject somewhat neglected in most textbooks. Applications to physics and chemistry. Non-trivial exercises and complete bibliography. SS

Number Theory, T(17: 2), S, P*, L. Number Theory. Helmut Hasse. Trans: Horst Günter Zimmer. Grund. der math. Wissenschaften, B. 229. Springer-Verlag, 1980, xvii + 638 pp, $49. [ISBN: 0-387-08275-1] A corrected and enlarged translation of the Third Edition of this classic, which originally appeared in 1949. Chapter 16 has been largely rewritten in order to remove an error. Includes some new bibliographical references and updatings in the tables. CEC

Number Theory, S(16-17), L. Selected Topics in Number Theory. Hansraj Gupta. Abacus Pr, 1980, 394 pp, $55. [ISBN: 0-85626-177-7] An interesting and unique compilation of topics from elementary number theory: the standard themes, primitive roots, reciprocity, power residues, diophantine equations, as well as Stirling and Bernoulli numbers, permutations, and partitions. Many exercises included. SG

Number Theory, T(18: 1), S, P. Lecture Notes in Mathematics-785: Diophantine Approximation. Wolfgang M. Schmidt. Springer-Verlag, 1980, x + 299 pp, $19.50 (P). [ISBN: 0-387-09762-7] An amplified version of the author's 1970 notes on the subject. The approach is that which was initiated by Thue and further developed by Siegel and Roth. The main emphasis is on approximation to algebraic numbers. The pace is relatively leisurely. Includes a substantial list of references. CEC

Linear Algebra, T(17: 1), S, P, L. The Symmetric Eigenvalue Problem. Beresford N. Parlett. P-H, 1980, xix + 348 pp, $25. [ISBN: 0-13-880047-2] This book is devoted to eigenvalue computations for real symmetric matrices. Includes a complete discussion of the QR and QL algorithms for diagonalizing small tridiagonal matrices along with recent advances in dealing with large matrices. Includes exercises, a list of references and an annotated bibliography. CEC

Algebra, S(18), P. Base Change for GL(2). Robert P. Langlands. Annals of Math. Stud., No. 96. Princeton U Pr, 1980, vii + 236 pp, $17.50; $7 (P). A slightly edited and expanded version of lecture notes for a course at the Institute for Advanced Study in 1975. Deals with representations of Gl(2) with emphasis on "the functoriality of automorphic forms with respect to...the L-group." Topics include Artin L-functions, spherical functions, orbital integrals, lifting, and trace formulas. Bibliography, indexes. JS

Algebra, S(18), P. Commutative Algebra, Second Edition. Hideyuki Matsumura. Benjamin/Cummings, 1980, xv + 312 pp, $19.50 (P). [ISBN: 0-8053-7026-9] No major changes in the text from the First Edition (TR, April 1971; ER, February 1972) except for revision of Chapter 6 on M-regular Sequences and Cohen-Macaulay Rings. An appendix has been added to supply some previously missing proofs as well as some more recent work of Falting, Marot, and Kunz. JS

Algebra, S(18), P. Lecture Notes in Mathematics-806: Burnside Groups. Ed: J.L. Mennicke. Springer-Verlag, 1980, 274 pp, $16.80 (P). [ISBN: 0-387-10006-7] Papers from a 1977 workshop held at the University of Bielefeld whose main purpose was to survey the present knowledge on Burnside groups, in particular the work of Novikov-Adian. Contributors include Adian, Alford, Pietsch, Grunewald, Havas, Mennicke, Newman, Hermanns, Levin, and Rosenberger. Newman has added a problem list and bibliography. JS

Algebra, P. Algebraic Theory of Quadratic Forms: Generic Methods and Pfister Forms. Manfred Knebusch, Winfried Scharlau. Birkhäuser Boston, 1980, 44 pp, $4.80 (P). [ISBN: 3-7643-1206-8] A clearly written, self-contained and short introduction to the subject. Suitable for a faculty or graduate level seminar. SG

Algebra, P. Les Hièrarchilongueurs. Mostafa Guennoun. Pure and Appl. Math., No. 55. Queen's U, 1980, iv + 149 pp, (P). An exposition of the author's work on the general theory of length functions over a ring with values in a partially ordered semigroup. SG

Algebra, T(17-18: 1-3), S, L.** Basic Algebra II. Nathan Jacobson. Freeman, 1980, xix + 666 pp, $30. [ISBN: 0-7167-1079-X] Clear, thorough treatment of nearly all the topics appropriate to a first year graduate course, e.g., homological algebra, group representations, commutative ideal theory, valuation theory. Many interesting exercises. Very readable. JG

Calculus, T(13: 1, 2). Calculus: A Modeling Approach, Second Edition. Marvin L. Bittinger. A-W, 1980, xi + 544 pp, $17.95. [ISBN: 0-210-01247-2] Brief intuitive treatment. Informal in style. The inclusion of somewhat more difficult exercises, numerous additional examples in business, economics, biology and medicine, and some changes in content highlight this edition. One of the best of its genre. (First Edition, TR, June-July 1976.) JK

Real Analysis, T(17). Measure Theory. Donald L. Cohn. Birkhäuser, 1980, ix + 373 pp, $19.95. [ISBN: 3-7643-3003-1] Standard measure theory topics in first six chapters, followed by measures on locally compact spaces, Polish spaces and analytic sets, and Haar measure. Exercises follow each section. SS

Complex Analysis, P. Distribution of Zeros of Entire Functions, Revised Edition. B. Ja. Levin. Transl. Math. Mono., V. 5. AMS, 1980, xii + 523 pp, $24 (P). [ISBN: 0-8218-4505-5] Inaccuracies and misprints of the first edition are corrected. Some post-1956 results are stated without proof, others appear in the "Further Developments" appendix. Expanded bibliography. Paperback. Primarily the same, albeit corrected, classic. PDH

Differential Equations, P. Opera Matematica, Volume II. Grigore C. Moisil. Editura Academiei (Romania), 1980, 358 pp, Lei 30. A second volume of papers in differential equations (Volume I, TR, December 1976). JAS

Differential Equations, P. Nonlinear Partial Differential Equations in Engineering and Applied Science. Ed: Robert L. Sternberg, Anthony J. Kalinowski, John S. Papadakis. Lect. Notes in Pure and Appl. Math., V. 54. Dekker, 1980, xvi + 480 pp, $55 (P). [ISBN: 0-8247-6996-1] Twenty-eight papers from the Conference on Nonlinear Partial Differential Equations in Engineering and Applied Science, sponsored by the Office of Naval Research, and held at the University of Rhode Island in June, 1979. Focuses on a variety of topics of specialized, contemporary concern to mathematicians, physical and biological scientists, and engineers. JK

Differential Equations, P*. Differential Equations. Ed: Shair Ahmad, Marvin Keener, A.C. Lazer. Acad Pr, 1980, ix + 278 pp, $20. [ISBN: 0-12-045550-1] Proceedings of the Eighth Fall Conference on Differential Equations held at Oklahoma State University in October 1979. In this most recent conference there was no special emphasis on any particular area of differential equations or on geographical locations of the participants, as had been the case in the first seven. Reproduced from an immaculately prepared typewritten manuscript. JK

Differential Equations, T(15: 1). Introduction to Ordinary Differential Equations, Third Edition. Shepley L. Ross. Wiley, 1980, 503 pp, $18.95. [ISBN: 0-471-03295-6] Retains clear exposition of previous editions. Some new features are discussions of matrix multiplication and inversion, material on Laplace transforms of step, translated, and periodic functions, and a short appendix on determinants. (First Edition, TR, February 1967; Second Edition, TR, May 1975; ER, November 1976.) MB

Differential Equations, P. Lecture Notes in Physics-120: Nonlinear Evolution Equations and Dynamical Systems. Ed: M. Boiti, F. Pempinelli, G. Soliani. Springer-Verlag, 1980, vi + 368 pp, $25 (P). [ISBN: 0-387-09971-9] Expanded versions of the talks given at the meeting at the University of Lecce, Italy, June 20-23, 1979. JAS

Differential Equations, P. Dynamical Systems. John Guckenheimer, Jürgen Moser, Sheldon E. Newhouse. Progress in Math., No. 8. Birkhäuser Boston, 1980, viii + 289 pp, $14 (P). [ISBN: 3-7643-3024-4] Notes from three lecture series presented at the June 1978 Centro Internazionale Matematico Estivo (C.I.M.E.) in Bressanone, Italy on structurally stable systems, bifurcation theory, and integrable Hamiltonian systems. LAS

Differential Equations, P. On the Asymptotic Analysis of Large-Scale Ocean Circulation. W.P.M. de Ruijter. Math. Centre Tracts, No. 120. Math Centrum, 1980, iii + 116 pp, Dfl. 14 (P). [ISBN: 90-6196-193-9] This study concerns the solution of a boundary value problem for a transport stream function that describes steady-state circulation in the ocean. The equation involves a fourth order differential operator in two space variables. TRS

Numerical Analysis, P. Numerical Treatment of Integral Equations. Ed: J. Albrecht, L. Collatz. Int. Ser. Num. Math., V. 53. Birkhäuser Boston, 1980, 275 pp, $33 (P). [ISBN: 3-7643-1105-3] Papers from the workshop on the title topic held at Oberwolfach, November 18-24, 1979. JAS

Numerical Analysis, P. Bifurcation Problems and their Numerical Solution. Ed: H.D. Mittelmann, H. Weber. Int. Ser. Num. Math., V. 54. Birkhäuser Boston, 1980, 243 pp, $24 (P). [ISBN: 3-7643-1204-1] Proceedings of a conference held in January 1980 at the University of Dortmund. Includes a survey article on the subject. Some of the papers are in German. AO

Functional Analysis, P. Lecture Notes in Mathematics-796: Duality in Measure Theory. Corneliu Constantinescu. Springer-Verlag, 1980, iv + 197 pp, $14 (P). [ISBN: 0-387-09989-1]

Functional Analysis, P. Special Topics of Applied Mathematics: Functional Analysis, Numerical Analysis and Optimization. Ed: J. Frehse, D. Pallaschke, U. Trottenberg. North Holland, 1980, viii + 248 pp, $39. [ISBN: 0-444-86035-5] Proceedings of an international seminar held in October 1979 at the Gesellschaft für Mathematik und Datenverarbeitung in honor of the retirement of Professor Dr. Ing. Heinz Unger, founder of the GMD. JK

Functional Analysis, P. Lecture Notes in Mathematics-803: Dirichlet Integrals on Harmonic Spaces. Fumi-Yuki Maeda. Springer-Verlag, 1980, x + 180 pp, $11.80 (P). [ISBN: 0-387-09995-6] Beginning with the concept of a "gradient measure" on general harmonic spaces, the author introduces generalized Green's potentials on self-adjoint harmonic spaces and integral representations. An analogue of the Royden boundary theory is developed. TAV

Optimization, T*(16-17: 1, 2), P, L. Network Flow Programming. Paul A. Jensen, J. Wesley Barnes. Wiley, 1980, xv + 408 pp, $25.95. [ISBN: 0-471-04471-7] Exposition of a broad range of single commodity network flow problems and techniques of solution. Good exercises, especially those requiring network formulation. Historical perspective, with references at end of each chapter, plus extensive bibliography. Intended primarily for the practitioner. JG

Optimization, T(17: 1), P. Algorithms for Network Programming. Jeff L. Kennington, Richard V. Helgason. Wiley, 1980, xiii + 291 pp, $23.50. [ISBN: 0-471-06016-X] Network flow models discussed include minimal cost, multicommodity, convex cost, generalized network flow, and network flow with side constraints. Complete descriptions of the algorithms needed, together with the basic data structures. Good bibliography. JG

Optimization, P. Variational Inequalities and Complementarity Problems: Theory and Applications. Ed: R.W. Cottle, F. Giannessi, J-L. Lions. Wiley, 1980, xvii + 408 pp, $48.95. [ISBN: 0-471-27610-3] Proceedings of an International School of Mathematics devoted to variational inequalities and complementarity problems in mathematical physics and economics, which was held in June 1978 at the Ettore Majorana Center for Scientific Culture in Erice, Sicily. TRS

Optimization, P. Probleme de Optimizare si Algoritmi de Approximare a Solutiilor. Constantin Ilioi. Editura Academiei (Romania), 1980, 202 pp, Lei 11 (P). A unified view of the main algorithms for unconstrained and constrained optimization as well as for discrete approximation of continuous optimization problems. JAS

Optimization, P. Lecture Notes in Control and Information Sciences--22 & 23: Optimization Techniques. Ed: K. Iracki, K. Malanowski, S. Walukiewicz. Springer-Verlag, 1980, $34.90 (P) each. Part 1, xvi + 569 pp [ISBN: 0-387-10080-6]; Part 2, xv + 621 pp, $34.90 (P) each. [ISBN: 0-387-10081-4] Proceedings of the Ninth IFIP Conference on Optimization Techniques held at Warsaw, Poland, September 4-8, 1979. JAS

Optimization, S(16-17), P, L. Computer Optimization Techniques. William Conley. Petrocelli Book, 1980, xii + 266 pp, $24. [ISBN: 0-89433-111-6] Working with numerous nonlinear optimization problems known to present difficulties when analyzed by standard techniques, the author shows how computer search methods can be used to obtain satisfactory results; Monte Carlo techniques for large problems are discussed. AWR

Analysis, P. Spectral Theory of Functions and Operators. Ed: S.M. Nikol'skii. Proc. of Steklov Inst. of Math., No. 130. AMS, 1980, v + 233 pp, $40 (P). [ISBN: 0-8219-3030-9]

Analysis, P. Tauberian Theory and its Applications. A.G. Postnikov. Proc. of Steklov Inst. of Math., No. 144. AMS, 1979, v + 138 pp, $26 (P). [ISBN: 0-8218-3048-1] An inviting exposition of classical Tauberian theory in 29 brief sections. LAS

Analysis, P. Opera Matematica, Functii Poliarmonice. Miron Nicolescu. Editura Academiei (Romania), 1980, 354 pp, Lei 27.

Analysis, P. Lecture Notes in Mathematics-814: Séminaire de Théorie du Potentiel Paris, No. 5. M. Brelot, G. Choquet, J. Deny. Springer-Verlag, 1980, iv + 239 pp, $14 (P). [ISBN: 0-387-10025-3]

Analysis, P. Lecture Notes in Mathematics-809: Recurrences and Discrete Dynamic Systems. Igor Gumowski, Christian Mira. Springer-Verlag, 1980, vi + 272 pp, $16.80 (P). [ISBN: 0-387-10017-2] Intended to be accessible to readers from diverse scientific disciplines who face time-evolving phenomena, this monograph provides a systematic and unified treatment of the important properties of first and second order recurrences. Includes substantial chapters on stochasticity in conservative, almost conservative, and strongly nonconservative recurrences. TRS

Differential Geometry, T(17-18: 1), L. Introduction to Global Analysis. Donald W. Kahn. Pure and Appl. Math., V. 91. Acad Pr, 1980, ix + 336 pp, $34.50. [ISBN: 0-12-394050-8] This text, based on a graduate course given at the University of Minnesota, gives a coherent introductory view of those aspects of differential geometry which are showing up in applications. In particular, the reader is introduced to Morse theory, Lie groups, dynamical systems, and catastrophe theory. 10 of the 11 chapters end with a short but substantial set of "problems and projects." Good bibliography, short index. JAS

Differential Geometry, P. Vector Bundles and Differential Equations. Ed: André Hirschowitz. Progress in Math., No. 7. Birkhäuser Boston, 1980, v + 249 pp, $14 (P). [ISBN: 3-7643-3022-8] Eight lectures from the Journées Mathématiques sur les Fibrés vectoriels et Equations différentielles held in Nice, France, June 12-17, 1979. JAS

Algebraic Geometry, P. Lecture Notes in Mathematics-813: Algebroid Curves in Positive Characteristic. Antonio Campillo. Springer-Verlag, 1980, v + 168 pp, $11.80 (P). [ISBN: 0-387-10022-9] Development of the theory of equisingularity of irreducible algebroid curves over an algebraically closed field of arbitrary characteristic, using the Hamburger-Noether expansion. JG

Geometry, T(18: 1), S, P. Translation Planes. Heinz Lüneburg. Springer-Verlag, 1980, ix + 278 pp, $29.80. [ISBN: 0-387-09614-0] This book focuses on the great progress of the last two decades in the theory of finite translation planes. Presupposes considerable knowledge of projective planes and group theory. A significant list of references. No exercises. CEC

Topology, P. Topology. Ed: A. Császár. North-Holland, 1980. Volume I, 651 pp; Volume II, 605 pp, $161 set. [ISBN: 963-8021-30-6] Detailed versions of 112 papers on a wide range of topics from the

Colloquium on Topology held in Budapest, August 7-11, 1978. JAS

Probability, T(18: 1), P*. Probability Inequalities in Multivariate Distributions. Y.L. Tong. Prob. and Math. Stat. Acad Pr, 1980, xiii + 239 pp, $29.50. [ISBN: 0-12-694950-6] Comprehensive, systematic, and essentially self-contained treatment. Includes applications to the areas of simultaneous confidence regions, hypothesis testing, multiple decision problems, and reliability theory. Extensive bibliography. RSK

Probability, P. Multicomponent Random Systems. Ed: R.L. Dobrushin, Ya. G. Sinai. Dekker, 1980, xi + 606 pp, $65. [ISBN: 0-8247-6831-0] This is Volume 6 of Advances in Probability and Related Topics. All the articles are translated from the work of mathematicians connected with the Institute for Information Transmission Problems in Moscow. JAS

Probability, P. Stochastic Filtering Theory. Gopinath Kallianpur. Appl. of Math., No. 13. Springer-Verlag, 1980, xvi + 316 pp, $29.80. [ISBN: 0-387-90445-X] Filtering theory is the heart of Ito's stochastic calculus. Assuming familiarity with the Wiener process, the author attempts to present the major results of Ito's work along with their application to stochastic integral representations. Includes a treatment of Gaussian stochastic equations and their role in linear filtering theory. Contains a useful bibliography. TAV

Probability, T(18: 1), S. The Central Limit Theorem for Real and Banach Valued Random Variables. Aloisio Araujo, Evarist Giné. Wiley, 1980, xiv + 233 pp, $26.95. [ISBN: 0-471-05304-X] Any book that begins with four pages of symbols and abbreviations calls for careful reading. The authors begin with weak convergence results for measure on metric space and proceed to develop a treatment of the CLT on a line that extends to Banach space valued random variables. Tough reading for any but an expert. Contains an extensive bibliography. TAV

Probability, P. Lecture Notes in Mathematics-816: Local Operators and Markov Processes. Lucretiu Stoica. Springer-Verlag, 1980, viii + 104 pp, $9.80 (P). [ISBN: 0-387-10028-8] Classical examples of the axiomatic potential theory of continuous Markov processes are associated with second order elliptic or hypoelliptic differential operators. Local operators on locally compact spaces provide analogous behavior and are studied by the author. A very abstract treatment. TAV

Probability, T(18: 2), P. Theorie zufälliger Prozesse. Alexander D. Wentzell. Math. Reihe, B. 65. Birkhäuser Boston, 1979, x + 253 pp, $44. [ISBN: 3-7643-1021-9] A translation, with corrections and minor additions, of the advanced Russian text published in 1975. Many problems, sketches of solutions for most of them. JD-B

Statistics, T(15-17: 1), S, P, L. Regression Analysis and Its Application: A Data-Oriented Approach. Richard F. Gunst, Robert L. Mason. Statistics, V. 34. Dekker, 1980, xiv + 402 pp, $39.75. [ISBN: 0-8247-6993-7] Emphasizes the problems of dealing with real data. Exercises and examples use real data. In addition to the central discussion of regression analysis, it provides a review of matrix algebra and discussions of variable selection techniques, multicollinearity effects, and biased estimates (e.g., ridge regression). FLW

Statistics, T(15-16: 1-3), S. Probability and Statistics in Engineering and Management Science, Second Edition. William W. Hines, Douglas C. Montgomery. Wiley, 1980, xvi + 634 pp, $22.95. [ISBN: 0-471-04759-71] Presupposes calculus and (in one chapter) matrix algebra. They use topics in probability and statistics plus multiple regression, design of experiments, quality control, Markov chains, queueing theory, and decision theory. (First Edition, TR, August-September 1973.) FLW

Statistics, T(16-18: 1, 2), S, P, L. Statistical Decision Theory: Foundations, Concepts, and Methods. James O. Berger. Series in Statistics. Springer-Verlag, 1980, xv + 425 pp, $24. [ISBN: 0-387-90471-9] Bayesian analysis, minimax analysis, invariance, sequential analysis, and complete classes. Presupposes calculus and some statistics. FLW

Statistics, S(14-18), P. Tables for Normal Tolerance Limits, Sampling Plans, and Screening. Robert E. Odeh, D.B. Owen. Statistics, V. 32. Dekker, 1980, ix + 316 pp, $49.75. [ISBN: 0-8247-6944-9]

Statistics, T(16-18: 1, 2), S, P, L. Statistical Methods for Comparative Studies: Techniques for Bias Reduction. Sharon Anderson, et al. Wiley, 1980, xiii + 289 pp, $24.95. [ISBN: 0-471-04838-0] Old and new methods for evaluating the effectiveness of new programs involving human populations. The general problem, matching, standardization and stratification, analysis of covariance, logit analysis, log-linear analysis, survival analysis, and repeated measure designs. No exercises. FLW

Statistics, P. Lecture Notes in Statistics-3: Benefit-Cost Analysis of Data Used to Allocate Funds. Bruce D. Spencer. Springer-Verlag, 1980, viii + 296 pp, $16.80 (P). [ISBN: 0-387-90511-1] An intensive study, using 1970 census data, of the effect of undercounting in non-uniformly distributed segments of the population. Since federal allocation of funds is based on this data, a benefit-cost analysis of collecting better data is undertaken. AWR

Statistics, S(18), P. Modern Statistics: Methods and Applications. Ed: Robert V. Hogg. Proc. of Symp. in Appl. Math., V. 23. AMS, 1980, vi + 110 pp, $18. [ISBN: 0-8218-0023-X] Lecture notes from the January 1980 AMS Short Course in San Antonio: samples and surveys, analysis of variance,

nonparametric tests, ordered parameters, time series. LAS

Statistics, P. Multivariate Statistical Analysis. Ed: R.P. Gupta. North-Holland, 1980, viii + 289 pp, $41.50. [ISBN: 0-444-86019-3] Proceedings of the Research Seminar at Dalhousie University, Halifax, held in October, 1979. Includes twenty-five papers from the conference, which emphasized "both the theory and applications of statistical distributions with reference to multi-dimensional random variables." RSK

Statistics, P. The Distribution of the Size of the Maximum Cluster of Points on a Line. Norman D. Neff, Joseph I. Naus. Selected Tables in Math. Stat., V. 6. AMS, 1980, vii + 207 pp, $12.80. [ISBN: 0-8218-1906-2] Tables give probabilities for the size of the largest cluster of random points on a subinterval of length p when N points are chosen independently from the uniform distribution on the unit interval, and the expected probabilities when N is a Poisson random variate. Includes a variety of applications. RSK

Computer Programming, T(13: 1), S. Structured BASIC and Beyond. Wayne Amsbury. Computer Sci Pr, 1980, xvii + 310 pp, $10.95 (P). [ISBN: 0-914894-16-1] Introductory text for those who favor a very structured approach to Basic programming. A structured pseudocode is used in place of flow-charts. Clear presentations. Ample exercises. With chapters on stacks and queues, linked lists and trees. JL

Computer Programming, T(13: 1). Elementary Computer Programming in Fortran IV, Second Edition. Boris W. Boguslavsky. Reston Pub, 1980, xi + 482 pp, $17.95 (P). [ISBN: 0-8359-1648-0] No mention of ANSI Fortran 77. Straightforward presentation of Fortran IV with answers to all odd problems in back. (First Edition, TR, June-July 1974.) LLK

Computer Programming, T(13: 1). COBOL Programming, A Structured Approach. Peter Abel. Reston Pub, 1980, xv + 408 pp, $12.95 (P). [ISBN: 0-8359-0833-X] Introductory text. DOS and OS. Emphases on program logic and debugging techniques. Many worked-out examples. JL

Computer Science, S(13-16). Minicomputer Systems. M. Vardell Lines. Winthrop Pub, 1980, x + 217 pp, $19.95. [ISBN: 0-87626-582-4] A nontechnical overview of the organizational and operational features of minicomputer systems, with relation to computer networks and systems design. For readers with some computer background. JL

Computer Science, S(14-16). Best of Interface Age, Volume 2: General Purpose Software. Ed: Interface Age Staff. Dilithium Pr, 1980, viii + 204 pp, $8.95 (P). [ISBN: 0-918398-37-1] Second of a five volume series. Contains "thirteen of the most-asked-for system and application software articles printed in Interface Age" magazine, plus an interesting article on the ASCII data alphabet. RSK

Computer Science, P. Lecture Notes in Computer Science-81: Data Base Techniques for Pictorial Applications. Ed: A. Blaser. Springer-Verlag, 1980, xi + 599 pp, $34.90 (P). [ISBN: 0-387-09763-5] Proceedings of an IBM-sponsored conference held in Florence, Italy, June 20-22, 1979. The papers are mostly expository, philosophy rather than code for algorithms, and cover geographic applications, geometric questions, image processing, and data base research. JAS

Operations Research, S(17-18), P. Studies on Mathematical Programming. Ed: A. Prèkopa. Akadèmiai Kiadó, 1980, 200 pp, $16. [ISBN: 963-05-1854-6] Collected papers from the Third Matrafured Conference of the Hungarian Academy of Science. A few of the 15 papers are surveys of particular theories; others deal with linear, nonlinear, stochastic, and parametric programming. AWR

Applications (Control Theory), P. Lecture Notes in Control and Information Sciences-24: Methods and Applications in Adaptive Control. Ed: H. Unbehauen. Springer-Verlag, 1980, vi + 309 pp, $21.60 (P). [ISBN: 0-387-10226-4] Proceedings of the symposium held at the Ruhr-University Bochum on March 20-21, 1980. This symposium brought together theoreticians possessing a solution to the stability problem and engineers with modern microprocessors. "It seems that today adaptive control is about ready for industrial applications." JAS

Applications (Cybernetics), P. Image Pattern Recognition. V.A. Kovalevsky. Trans: Arthur Brown. Springer-Verlag, 1980, xi + 241 pp, $29.80. [ISBN: 0-387-90440-9] Description of the recognition problem. Solutions of several image recognition problems using parametric models of the pattern generating process. Description of the Cars character reader. JG

Applications (Ecology), P*. Statistical Ecology Series, V. 4-13. Ed: G.P. Patil. Intern Co-op Pub, 1979. [ISBN: 0-89974-000-6] Statistical Distributions in Ecological Work. Ed: J.K. Ord, G.P. Patil, C. Taillie. V. 4. xxx + 464 pp, $40 [ISBN: 0-89974-001-4]; Sampling Biological Populations. Ed: Richard M. Cormack, Ganapati P. Patil, Douglas S. Robson. V. 5. xxviii + 392 pp, $40 [ISBN: 0-89974-002-2]; Ecological Diversity in Theory and Practice. Ed: J. Frederick Grassle, et al. V. 6. xxxi + 365 pp, $40 [ISBN: 0-89974-003-0]; Multivariate Methods in Ecological Work. Ed: L. Orloci, C.R. Rao, W.M. Stiteler. V. 7.; Spatial and Temporal Analysis in Ecology. Ed: R.M. Cormack, J.K. Ord. V. 8. xxix + 356 pp, $40 [ISBN: 0-89974-005-7]; Systems Analysis of Ecosystems. Ed: G.S. Innis, R.V. O'Neill. V. 9. xxvii + 402 pp, $40 [ISBN: 0-89974-006-5]; Compartmental Analysis of Ecosystem Models. Ed: James H. Matis, Bernard C. Patten, Gary C. White. V. 10. xxx + 368 pp, $45 [ISBN: 0-

89974-007-3]; Environmental Biomonitoring, Assessment, Prediction, and Management--Certain Case Studies and Related Quantitative Issues. Ed: John Cairns, Jr., Ganapati P. Patil, William E. Waters. V. 11. xxviii + 438 pp, $45 [ISBN: 0-89974-008-1]; Contemporary Quantitative Ecology and Related Ecometrics. Ed: Ganapati P. Patil, Michael L. Rosenzweig. V. 12. xxxii + 695 pp, $60 [ISBN: 0-89974-009-X]; Quantitative Population Dynamics. Ed: D.G. Chapman, V. Gallucci. V. 13. Edited research papers and research-review-expositions from a Satellite Program in Statistical Ecology during 1977 and 1978. The Program consisted of NATO Advanced Study Institutes at College Station in Texas, Berkeley in California, and Parma in Italy; a NATO Advanced Research Institute at Parma; International Statistical Ecology Program Research Conferences, Seminars, and Workshops at College Station, Berkeley, Parma, and Jerusalem; and a Research Conference at Jerusalem. The emphasis was on research, review and exposition concerned with the interface between quantitative ecology and relevant quantitative methods. A "state-of-the-art" series. RSK

Applications (Engineering), P. New Approaches to Nonlinear Problems in Dynamics. Ed: Philip J. Holmes. SIAM, 1980, xii + 529 pp, $42.50. [ISBN: 0-89871-167-3] Proceedings of an Engineering Foundation Conference held December 9-14, 1979, at the Asilomar Conference Grounds. Papers given in the areas of mathematical methods, aerospace and mechanical engineering, chemical engineering, electrical and civil engineering, bifurcation with symmetry, stochastic problems, strange attractors, large scale and distributed systems. TRS

Applications (Management), P. The Manager's Guide to Statistics and Quantitative Methods. Donald W. Kroeber, R. Lawrence LaForge. McGraw, 1980, xii + 348 pp, $18.95. [ISBN: 0-07-035520-7] The premise of the authors is that managers must be able to understand what their technical personnel are talking about. With this in mind, they describe the significance of several statistical and quantitative (linear programming, queueing, inventory) methods clearly and without fuss. If every manager had such a book more sense would be made of the work of trained technical people--a worthwhile effort. TAV

Applications (Modeling), T(16-18: 1, 2), S, L. Mathematical Modeling with Computers. Samuel L.S. Jacoby, Janusz S. Kowalik. P-H, 1980, xi + 292 pp, $21.50. [ISBN: 0-13-561555-0] A general guide for mathematical modeling with computers, including discussions on each step of the process: modeling objectives, method identification, analysis, computer implementation, model validation and interpretation. No exercises. LCL

Applications (Physics), P. Bifurcation Phenomena in Mathematical Physics and Related Topics. Ed: Claude Bardos, Daniel Bessis. Reidel, 1980, ix + 596 pp, $59.50. [ISBN: 90-277-1086-4] Proceedings of the NATO Advanced Study Institute held at Cargèse, Corsica, France, June 24-July 7, 1979. Papers cover a wide range, from algebraic geometry to the origin of ice ages on earth. JAS

Applications (Physics), P. Lecture Notes in Physics-116: Mathematical Problems in Theoretical Physics. Ed: K. Osterwalder. Springer-Verlag, 1980, viii + 412 pp, $27.70 (P). [ISBN: 0-387-09964-6] The proceedings of the International Conference on Mathematical Physics held at the Swiss Federal Institute of Technology in Lausanne, August 20-26, 1979. (I.M. Singer's main lecture is not included, cf. Singer's contribution to the Proceedings of the 1979 Cargèse summer school on Recent Developments in Gauge Theories.) This continues the tradition of meetings held in Moscow (1972), Warsaw (1974), Kyoto (1975), Rome (1977). The next conference is scheduled for Berlin in 1981. JAS

Applications (Physics), P. Lecture Notes in Physics-118: Quantum Chromodynamics. Ed: J.L. Alonso, R. Tarrach. Springer-Verlag, 1980, ix + 424 pp, $27.70 (P). [ISBN: 0-387-09969-7] Six lengthy research survey lectures on hadrons, nonabelian gauge theory, and quantum chromodynamics from the Tenth International Seminar sponsored by the Spanish Interuniversity Group of Theoretical Physics, held at Jaca, Huesca, Spain, in June 1979. Quantum chromodynamics does for the weak interactions (involving hadrons, leptons, quarks and gluons) what quantum electrodynamics does for electromagnetic interactions. ("Chromo" derives from the convention of naming quark degrees of freedom with colors rather than with numbers.) LAS

Applications (Physics), S(18), P. Introduction to the Theory of Quantized Fields, Third Edition. N.N. Bogoliubov, D.V. Shirkov. Trans: Seweryn Chomet. Wiley, 1980, xv + 620 pp, $55. [ISBN: 0-471-04223-4] Translation of new edition of classic monograph on quantum field theory. Chapters on functional integration and the renormalization group have been rewritten. AO

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; WC: William Carlson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Oste[illegible], St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schu[illegible] Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

Simeon Reich, Professor at the University of Southern California, will be Visiting Professor at the University of California, Berkeley, for the spring quarter 1981.

Professor *Paul Chernoff* will be on sabbatical leave from the University of California, Berkeley, for the winter and spring quarters, 1981.

Dr. *Frederick Steen* has retired from the staff of Allegheny College and is now Professor Emeritus of Mathematics.

At the University of British Columbia Associate Professor *Frank Herbert Clarke* has been promoted to Professor. He has been the holder of a Canada Council Killam Fellowship in 1979/80 and again this year. *David Boyd*, Professor of Mathematics at the University of British Columbia, recently shared the Steacie Prize.

At the University of Houston, Central Campus, Associate Professors *Charles P. Benner* and *Homer F. Walker* have recently been promoted to Professor.

James C. Reber, Chairman of the Department of Mathematics at Indiana University of Pennsylvania, has been promoted from Associate Professor to Professor.

Bernard Vinograde has retired from the Department of Mathematics at Iowa State University with title of Professor Emeritus.

The following promotions have been announced at Iowa State University: *Gary M. Lieberman* from Instructor to Assistant Professor; Assistant Professor *Harry F. Smith* to Associate Professor; and Associate Professor *Don L. Pigozzi* to Professor.

Stuart Goff, Associate Professor of Mathematics at Keene State College, has been appointed Dean of the Division of Sciences.

At the University of Louisville, *Michael Jacobson*, formerly at Emory University, has been appointed Assistant Professor of Mathematics. Professor *Thomas M. Jenkins* has taken a temporary position as Assistant Director of Academic Computing.

Dr. *William V. Smith*, formerly at Texas Technological University, has been appointed Assistant Professor of Mathematics at the University of Mississippi.

Joseph Lehner of the University of Pittsburgh has retired with the rank of Professor Emeritus.

At Salem College, Assistant Professor *Lewis Lum* has been promoted to Associate Professor and named chairman of the Department of Mathematics. Associate Professor *David C. Kurtz* is on leave-of-absence, 1980-82, to serve on the faculty of Chancellor College, Zomba, Malawi.

At San Antonio College Dr. *Frederick Stiles* of the University of Texas at Austin has been appointed Instructor in Mathematics. Associate Professor *Raymond Tebbetts* has been promoted to Professor and Instructor *Suren K. Shrivastava* has been promoted to Assistant Professor.

William O. McClung, formerly Assistant Professor of Mathematics at San Bernardino State College, has been appointed Assistant Professor of Mathematics at the University of San Francisco.

At the University of South Carolina, Professor *G. Lorentz* of the University of Texas has been named Visiting Professor. Associate Professors *H.E. Scheiblick*, *R.L. Taylor* and *William T. Trotter, Jr.*, have been promoted to Professor. Professor *T.H. Lee* has retired. Professor *S. Riemenschneider* of the University of Alberta has been appointed Visiting Associate Professor.

Dr. *Hueytzen James Wu* of the University of Arkansas has been appointed Assistant Professor of Mathematics at Texas Agricultural and Industrial University.

Dr. *Kevin Andrews*, formerly of Catholic University, has been appointed Assistant Professor of Mathematics at Texas Agricultural and Mechanical University.

Associate Professor *William R. Nico* of Tulane University has been promoted to Professor.

At the Worcester Polytechnic Institute, *M.L. Gardner*, formerly of North Carolina State University has been appointed Assistant Professor in the Department of Mathematical Sciences. Associate Professor *W.B. Miller* has been promoted to Professor.

James De Franza, formerly at St. Lawrence University, has been appointed Instructor in the Department of Mathematical and Computer Sciences at Youngstown State University.

Dr. *Theophil H. Hildebrandt*, Professor Emeritus at the University of Michigan, died October 9, 1980, at the age of 92. He was a member of the Association for 64 years.

Professor *Einar Hille* of the University of California at San Diego died February 12, 1980. A world-famous mathematician, he was a member of the Association for 57 years. A fuller account of his work will appear later in this MONTHLY.

Dr. *C.D. Olds*, Professor Emeritus at San Jose State University, died November 11, 1979. He was a member of the Association for 43 years.

Dr. *J. William Peters* of Champaign, Illinois, died February 28, 1979, at the age of 76. He was a member of the Association for 53 years.

Professor *Charles J. Pipes* of Southern Methodist University, died August 18, 1980. He was a member of the Association for 33 years.

Mr. *Paul H. Renton* of Westbrook, Connecticut, died August 12, 1980. He was a member of the Association for 30 years.

ARTHUR SARD

Arthur Sard, whose death was briefly noted in this MONTHLY for November 1980, was born in New York City in 1909. He received his doctorate in 1936 at Harvard and joined the faculty of Queens College, CUNY, in 1937. He continued in that association all his life, becoming Professor Emeritus in 1971. From 1943 to 1946 he was a Research Associate with the Division of War Research of the Columbia University Applied Mathematics Group. He was a Research Associate with the Department of Mathematics at the University of California at La Jolla from 1971 to 1975. From 1978 to 1979 he was associated with a research institute at the Gesamthochschule Siegen in West Germany.

Dr. Sard was the author of "Linear Approximation," published in 1963 by the American Mathematical Society and, with Sol Weiñtraub, of "A Book of Splines," published in 1971 by John Wiley and Sons. Nearly forty of his papers appeared in many prestigious journals during the period extending from 1938 to the present. His work in approximation theory and related topics was a significant and memorable contribution to the development of the subject.

OLYMPIAD NEWS

The number of nations interested in mathematics contests grows apace. The Brazilian Mathematical Society plans to sponsor the First Pan-American Mathematical Olympiad, probably in early 1982. It will be held in São Paulo or Rio de Janeiro. Countries interested in participating should contact Professor João Bosco Pitombeira, Departamento de Matematica, Pontificia Universidade Católica de Rio de Janeiro, Rua Marquês de São Vicente 225, Rio de Janeiro - RJ - Brazil, 22 453.

REMEDIAL-DEVELOPMENTAL CONFERENCE

City University of New York through the Instructional Resource Center and in cooperation with Networks and the CUNY Mathematics Discussion Group plans a national conference on REMEDIAL AND DEVELOPMENTAL MATHEMATICS IN COLLEGE: ISSUES AND INNOVATIONS which will take place April 9-11, 1981, at the Hotel Roosevelt in New York City. For information and registration forms please write to Dr. *Goeffrey Akst*, Conference Chair, CUNY Instructional Resource Center, Box M, 535 East 80th Street, New York, N.Y. 10021.

TABS PROJECT AT OHIO STATE UNIVERSITY

The College of Education of The Ohio State University has begun a project to develop and disseminate exemplary curricular materials in which high technology is used to teach basic mathematical skills including problem solving, estimation, computer literacy, etc. Funded by the U.S. Department of Education, the project will collect and evaluate existing educational software for mini-computers like APPLE, TRS-80, and PET. Other curricular elements will be developed by the project under the direction of Suzanne Damarin, Marlin Languis, and Richard Shumway. The curricula will be field tested and disseminated nationally.

Individuals or groups who have developed programs related to mathematics at the upper elementary school level are invited to submit them for possible inclusion for national dissemination. To have materials considered, send a cassette tape or floppy disk together with a printout, machine documentation and any related information to Dr. Suzanne K. Damarin, TABS Project, Arps Hall 202-A, 1945 N. High St., Columbus, Ohio 43210. For further information write or call Dr. Damarin at 614-422-1527.

COMING SOON IN THIS MONTHLY

The following articles will appear in the AMERICAN MATHEMATICAL MONTHLY for February 1981:

Thomas Zaslavsky, *The Geometry of Root Systems and Signed Graphs*
Chester J. Salwach, *Planes, Biplanes, and Their Codes*
Hans Willi Siegberg, *Some Historial Remarks Concerning Degree Theory*

The following articles are among those which the editors have accepted for later issues of the MONTHLY. The order of listing does not indicate the order in which they will appear.

William Abikoff, *The Uniformization Theorem*
Ray Bobo, *Foursomes, Fivesomes, and Orgies*
Ezra A. Brown, *The First Proof of the Quadratic Reciprocity Law, Revisited*
Philip J. Davis, *Are There Coincidences in Mathematics?*
Solomon W. Golomb, *Irrational Sums and Twin Primes*
H.B. Griffiths, *Cayley's Version of the Resultant of Two Polynomials*
Frank D. Grosshans, *Rigid Motions of Conics*
R. Arthur Knoebel, *Exponentials Reiterated*
Russell Merris, *Pólya's Counting Theorem via Tensors*
Louis J. Ratliff, Jr., *A Brief History and Survey of the Catenary Chain Conjectures*
John L. Troutman and W. Hrusa, *Elementary Characterizations of Classical Optima*

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

OCTOBER MEETING OF THE OHIO SECTION

The Ohio Section of MAA held its fall meeting at John Carroll University, University Heights, Ohio, October 17-18, 1980. One hundred and forty-five people registered for the meeting. Section Chairman C.A. Long presided; DJ. Horwath was the program chairman.

Invited addresses included: *Recent Implications of the Work of L.J. Rogers*, by G.E. Andrews, The Pennsylvania State University; and *An Introduction to Coding Theory*, by D.K. Ray-Chaudhuri, the Ohio State University

The following contributed papers were also presented:

Bayes' Discrete Data Analysis, J. Albert and A. Gupta, Bowling Green State University
A Report on (P and Q) Polynomial Association Schemes, E. Bannai, Ohio State University
Projective Concepts in Torsion Theories, A. Benander, John Carroll University
Finite Sigma-length in Torsion Theoreis, A. Benander, John Carroll University
Experiments in the Classroom--Using "Dots" to Determine Area, K.B. Cummins, Kent State University
A Model of Epidemics in Third World Children, P.J. Gingo, University of Akron
Regular Graphs with Small Excess, T. Ito, Ohio State University
A Finite Mathematics Problem Arising from Multiprocessing, Sr. T.M. McCloskey, Notre Dame College
Centers and Centroids in Graphs, Z. Miller, Miami University
Automorphism Groups of Graphs with Pendant Vertices and the Ulam Graph Reconstruction Problem, N. Robertson, Ohio State University
6-Flows on Graphs, P.D. Seymour, Ohio State University
Glimpses of Malawi, Africa, J.L. Smith, Muskingum College
Finite Optimization Algorithms in Operations Research and Graph Theory, D. Solow, Case Western Reserve University

The meeting agenda also included meetings of The Executive Committee and of ad hoc committees--Committee on Curriculum, Committee on Section Activity, and Committee on Teacher Training and Certification. Breakfast meetings were also held by the MAA Ohio Section campus representatives, and by The Association of Women in Mathematics.

Meeting highlights included discussion sessions and special presentations. *A Panel Discussion: The Computer Science Curriculum* was conducted by Z.A. Karian, Denison University (moderator); L. Benners, Western Electric Corp., Columbus; and L. White, Ohio State University. *A Special Session: Algebraic Graph Theory* was moderated by N. Robertson, Ohio State University. *"Swap" Session* presentations included: *Ways of Attracting High School Students to the Study of Mathematics*, moderated by J.A. Engle, Ohio State University; *Pre-Calculus Mathematics*, moderated by H.L. Putt, University of Akron; and *Video Cassettes for Supplementary Instruction*, Moderated by C.P. Yang, Miami University, Middletown. Also, a special Biographies-of-Great-Mathematicians lecture was presented: *Lewis Carroll*, by R.J. Kolesar, John Carroll University.

The officership and committee chairmanships for academic year 1980-81 include: *Executive Committe*--C.A. Long (Bowling Green State University), Section Chairman; D.O. Koehler (Miami University) Section Past-Chairman; G. Mavrigian (Youngstown State University), Secretary-Treasurer; S.W. Hahn (Wittenberg University), Sectional Governor; D.J. Horwath (John Carroll University), Program Committee Chairman; and J.D. Faires (Youngstown State University). Program Committee--D.J. Horwath, Chairman; A.G. Poorman (Ashland College); and J.P. Leitzel (Ohio State University). *Ad Hoc Committee Chairmen*-Committee on Curriculum: H.L. Putt (University of Akron). Committee on Computing: Z.A. Karian (Denison University). Committee on Section Activity: P.H. Schmidt (University of Akron). Committee on Teacher Training and Certification: W.A. Kirby (Bowling Green State University). Publications Officer and Newsletter Editor: R.A. Little (Baldwin-Wallace College). Representative to the Two-Year College Mathematics Journal: C.P. Yang (Miami University, Middletown). High School Mathematics Competition Supervisor: L.J. Schneider (John Carroll University).

Professor L.J. Schneider was in charge of meeting arrangements for the host institution.

OCTOBER MEETING OF THE INDIANA SECTION

The fall meeting of the Indiana Section of the MAA was held at DePauw University on Saturday, October 18, 1980, with 60 members present. The chairman of the Section, Duane Deal of Ball State University, presided.

The following papers were presented:

The Geometry of Vision and Visual Illusions, Wayne M. Zage, Ball State University
How Napier Computed His Table of Logarithms, Meyer Jerison, Purdue University
Can a Mathematician Find Happiness in the Computing Field? William A. Marion, Valparaiso University
Computer Graphics: A Rpview of the Present State of the Art, Mario Borelli, Notre Dame University
The Hudson-Nash Merger and the Equations of Thermodynamics, Dennis G. Collins, Valparaiso University
Crime by Computer, Michael C. Gemignani, Indiana University-Purdue University at Indianapolis

CALENDAR OF FUTURE MEETINGS

Sixty-fourth Annual Meeting, San Francisco, California, January 9–11, 1981.
Sixty-first Summer Meeting, Pittsburgh, Pennsylvania, August 17–19, 1981.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Duquesne University, Pittsburgh, Pennsylvania, May 15–16, 1981.
EASTERN PENNSYLVANIA AND DELAWARE, Saturday before Thanksgiving.
FLORIDA, Bethune Cookman College, Daytona Beach, March 6–7, 1981.
ILLINOIS, Illinois State University, Normal, May 1–2, 1981.
INDIANA, Indiana University–Purdue University, Indianapolis, April 11, 1981.
INTERMOUNTAIN, Brigham Young University, Provo, Utah, April 10–11, 1981.
IOWA, Coe College, Cedar Rapids, April 10–11, 1981 (tentative date).
KANSAS, Benedictine College, Atchison, April 11-12, 1981.
KENTUCKY, Jefferson Community College, Louisville, April 3–4, 1981.
LOUISIANA-MISSISSIPPI, Mississippi State University, Mississippi State, February 13–14, 1981.
MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA, William and Mary College, Williamsburg, Virginia, April 11, 1981 (tentative date).
METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.
MICHIGAN, Oakland University, Rochester, May 1–2, 1981.
MISSOURI, Northwest Missouri State University, Maryville, April 10–11, 1981.
NEBRASKA, University of South Dakota, Vermillion, April 10–11, 1981.
NEW JERSEY, Seton Hall University, South Orange, spring 1981.
NORTH CENTRAL, Mankato State University, Mankato, Minnesota, May 1-2, 1981.
NORTHEASTERN, Saturday before Thanksgiving and third week in June.
NORTHERN CALIFORNIA, California State University, Hayward, March 1981 (tentative date).
OHIO, Miami State University, Oxford, April 10-11, 1981.
OKLAHOMA-ARKANSAS, Oklahoma Christian College, Oklahoma City, March 27–28, 1981.
PACIFIC NORTHWEST, second Saturday in June. Deadline for papers six weeks before meeting.
ROCKY MOUNTAIN, Colorado College, Colorado Springs, May 1–2, 1981.
SEAWAY, first Saturday in November and Saturday in late April. Deadline for papers six weeks before meeting.
SOUTHEASTERN, University of Alabama, Birmingham, April 10–11, 1981.
SOUTHERN CALIFORNIA, first or second Saturday in March.
SOUTHWESTERN, New Mexico State University, Las Cruces, April 1981.
TEXAS, San Antonio College, San Antonio, April 10–11, 1981.
WISCONSIN, University of Wisconsin, La Crosse, late March—early April 1981.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Toronto, January 3–8, 1981.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES
AMERICAN MATHEMATICAL SOCIETY, San Francisco, California, January 7–10, 1981.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.
ASSOCIATION FOR SYMBOLIC LOGIC, San Francisco, California, January 9–10, 1981.
ASSOCIATION FOR WOMEN IN MATHEMATICS, San Francisco, California, January 7–11, 1981.
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, St. Louis, Missouri, April 22–25, 1981.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Four Seasons Sheraton, Toronto, Canada, May 4–6, 1981.
PI MU EPSILON
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.

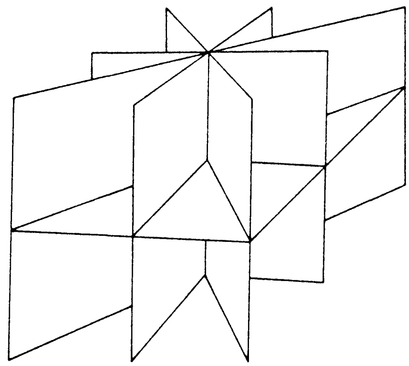
CONTENTS

FEBRUARY

THE AMERICAN MATHEMATICAL MONTHLY

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 88, Number 2



1981

Detailed contents on cover 4

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 2**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. (This is a change from the style previously used in this MONTHLY.) All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Research Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201; NOTES, etc.: to the corresponding Associate Editor; REPRINT PERMISSION: to R. P. BOAS (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N.Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are \$25 for each of the first two years of membership and \$30 thereafter. Student membership is available with annual dues of \$15. For nonmembers the subscription price is \$36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

AWARD FOR DISTINGUISHED SERVICE TO PROFESSOR RALPH P. BOAS, JR.

Ralph P. Boas, Jr., is a man who takes a constructive interest in everything that is going on around him in mathematics. This was the case when I first met him in 1934, and this has continued to be the case ever since. The result is that he has accomplished a great deal in many ways.

In 1934, Boas was a graduate student at Harvard; with Oxtoby and other fellow students, he pursued with enthusiasm the new ideas in real analysis stimulated by J. L. Walsh, D. V. Widder, and other faculty members. That same clear-headed and precise enthusiasm has grown and continued in many directions: in understanding polynomial expansions, the growth of entire functions, the integrability of trigonometric series, the managing of *Mathematical Reviews*, the editing of journals, the encouragement of graduate students, the guidance of the Committee on the Undergraduate Program in Mathematics, and now in editing this MONTHLY. Wherever he has turned, good results have turned up.

Boas was born in 1912 in Walla Walla, Washington. He came to Harvard University for his undergraduate work, leading to an A.B. degree in 1933 and a Ph.D. in 1934. He then won a National Research Fellowship for two years. This enabled him to learn more about the beauties of real analysis from Salomon Bochner at Princeton University and then from that grand master, G. H. Hardy, at Cambridge, England. From 1939 to 1945, Boas was engaged in teaching, first at Duke University, next at the Naval Preflight School in Chapel Hill, and then at Harvard, filling in for the regular faculty then on leave to wage the war. In 1945, his major contribution to *Mathematical Reviews* (*MR*) began. This was the time when he succeeded Willy Feller as Executive Editor of *MR*; in this way he became the first mathematician employed full-time by the American Mathematical Society. Recall that the *Reviews* had been started by the Society in 1940. They were located in Providence and guided by J. D. Tamarkin, Oswald Veblen, and Otto Neugebauer. The last had brought from Denmark his enormous insight and experience with the corresponding German journal *Zentralblatt*. When the process of reviewing mathematical articles became more extensive in 1945, Boas was the only editor; his subsequent work displayed the ability to cover every field and thus to make a vital contribution to the full development of this indispensable aid to mathematical research. His full-time work for *Mathematical Reviews* ended in 1950, but his interest did not end, as exhibited by his service on the editorial committee for *MR* (1954 to 1959) and on the Mathematical Reviews Crisis Committee in 1973—a committee which again restored the effective organization of this journal. His clear understanding of the importance of reviewing is well stated in his own article describing the distinguished service of Otto Neugebauer (this MONTHLY, vol. 86, February 1979, pp. 77/78).

In 1950, Boas became Professor of Mathematics at Northwestern University; he has continued until today in that role. At Northwestern he has served the university in many different ways. He was the Chairman of the Mathematics Department for fifteen years, and spokesman for a departmental committee for two additional years. He had at least nine Ph.D. students, and taught them and others with enthusiasm. He was active beyond the Mathematics Department; for example, he helped to pick three successive deans at Northwestern. One reason for picking three was the first choice did very well and so rapidly went on to become provost and then president at other universities.

Soon after coming to Northwestern, Ralph became active in the work of the Mathematical Association of America—in practically every aspect of this Association. He appears to have served on at least 25 different committees in a period of twenty years! As Chairman of Publications (1966 to 1967 and 1968 to 1969) he took a special interest in the Carus monograph series. Incidentally, that series published his enthusiastic book *Primer of Real Functions*, which was dedicated to "my epsilons." (There are three such, no longer epsilons. The youngest has just

RALPH P. BOAS, JR.

received a Ph.D. in Several Complex Variables.) But to return—that Carus volume of Boas, under the lucky number thirteen, managed to bring to its readers a clear and enticing understanding of many interesting results in real variables.

In 1957, Ralph announced his interest in good teaching of classical ideas by an article in this MONTHLY, "If this be treason . . ." (vol. 64, pp. 247–249). He took these interests quite seriously and soon had occasion to express them as one of the leading workers in that extraordinary effort of this Association, The Committee on the Undergraduate Program in Mathematics (CUPM). For this committee, he served as Chairman of the pregraduate panel from 1963 to 1965, and as Chairman of the committee itself from 1968 to 1970. In the first capacity, he helped produce the "pale green report," which described a realistic undergraduate program for students planning to go on to graduate school. In the second capacity, he guided the revision of the GCMC (the General Curriculum for Mathematics in Colleges).

We can take pride in his accomplishments and those of CUPM generally. Of the various college curriculum committees in different scientific fields, CUPM was the most effective—probably because it was not located at one college, but was responsible to a national association.

Space does not allow a list of all of the other effective things which Boas did for the Association. They did culminate in two high points: his service as President, 1973 to 1974, and his current work as editor of this MONTHLY, 1977 to 1981.

In the midst of all this activity, Ralph Boas has continued to be fascinated by trigonometric series, entire functions, and many other subtle phenomena of real variable theory. He has published about one hundred and fifty research articles. His collaborators number at least twenty and include his wife. His Erdős number is one (i.e., he has collaborated with Erdős). His effective knowledge of Russian has also been put to use in his translation of a number of Russian articles, as, for example, a recent important monograph by P. K. Suetin, "Polynomials Orthogonal over a Region and Bieberbach Polynomials" (*MR* 57, no. 3732b).

He has written two authoritative monographs for the *Ergebnisse* series, one in 1958 with R. C. Buck, "Polynomial Expansions of Analytic Functions," and another (on his own) in 1967, "Integrability Theorems for Trigonometric Transforms." There is one larger book, his 1954 volume, in which he pulled together many recent results (including his own) on the growth properties of entire functions. That splendid volume has been the bible in its subject ever since.

There have been many other Boas activities: A Trustee for the AMS; President of a chapter of AAUP; Chairman of Section A of AAAS; Chairman of the Mathematics Committee for the Advanced Graduate Record Exams (GRE); all told, his career has carried out his original enthusiasm for mathematics in all its aspects: research, monographs, books, teaching, editing, testing, managing, and organizing. It is the quality of his accomplishments and the breadth of his activities that render him today the appropriate candidate for the Distinguished Service Award of the Association.

Saunders Mac Lane

AWARD OF THE CHAUVENET PRIZE TO PROFESSOR KENNETH I. GROSS

The Board of Governors of the Mathematical Association of America voted to award the 1981 Chauvenet Prize to Professor Kenneth I. Gross for his paper "On the Evolution of Noncommutative Harmonic Analysis," which appeared in this MONTHLY, 85 (1978) 525–548.

A certificate and monetary award in the amount of five hundred dollars were presented to Professor Gross at the Business Meeting of the Association on January 7, 1981.

The Chauvenet Prize is awarded for a noteworthy paper of an expository or survey nature published in English that comes within the range of profitable reading for members of the Association. The purpose of the prize is to stimulate the writing of expository and survey articles. The 1981 Prize, awarded for a paper published in the three-year period 1977–79, is the twenty-ninth award of the Chauvenet Prize since its institution by the Association in 1924. For the names of the previous winners, see this MONTHLY, 71 (1964) 589; 72 (1965) 2–3; 74 (1967) 3; 75 (1968) 3–4; 77 (1970) 117–118; 78 (1971) 112–113; 79 (1972) 112–113; 80 (1973) 120; 81 (1974) 113–114; 82 (1975) 108–109; 83 (1976) 84–85; 84 (1977) 417; 85 (1978) 74–75; 86 (1979) 79; 87 (1980) 80. The award-winning papers for the years 1924–1976 are now available in the two-volume collection *The Chauvenet Papers*, published by the MAA.

Kenneth I. Gross was born on October 14, 1938, in Malden, Massachusetts, and was educated in the public schools of the nearby cities of Chelsea and Everett. His A.B., magna cum laude, with distinction in mathematics and physics, was received in 1960 from Brandeis University. Professor Gross also earned an M.A. from Brandeis University in 1962 and was granted a Ph.D. by Washington University in 1966.

From 1966 to 1973, Professor Gross served as Assistant Professor of Mathematics, first at Tulane University and then at Dartmouth College. In 1973, he joined the faculty of the University of North Carolina and has served as both Associate Professor and Professor of Mathematics. Professor Gross has held visiting positions at the University of California, Irvine (1972–73), the University of Utah (1976–77), and Academia Sinica, Taiwan (1978–79).

Professor Gross has participated in many aspects of MAA activities. In particular, he served as Chairman of the Arrangements Committee for the 1972 AMS-MAA Summer Meeting at Dartmouth College, as invited lecturer for the 1977 Annual Meeting held in St. Louis, as an invited lecturer for the Southeastern Section of the Association, and as recipient of the Lester R. Ford Award for outstanding expository articles published in this MONTHLY. The Lester R. Ford Award was in honor of the paper for which the 1981 Chauvenet Prize is granted.

The research interests of Professor Gross are in group representations and harmonic analysis, and he has published numerous articles in these areas. He also has in preparation the books *An Introduction to Representations of Groups and Analysis on Euclidean Spaces* and *A Primer on Vectors, Matrices, and Linear Algebra*, as well as his Taiwan lecture notes, *Group Representations and Harmonic Analysis on Homogeneous Spaces*.

Professor Gross has been the recipient of several fellowships and numerous National Science Foundation grants. He has been an active participant and organizer of conferences and is frequently an invited lecturer at departmental colloquia. He is also a frequently honored teacher and has been accorded highest ratings on undergraduate teaching at each of the institutions where he has been a faculty member.

Professor Gross is a member of the American Association of University Professors, the American Mathematical Society, and the Mathematical Association of America.

Upon learning of the selection of his paper for the Chauvenet Prize, Professor Gross expressed his sincere gratitude to the MAA for such a high honor. He added that he was deeply touched by this recognition of his efforts, which will serve as a stimulus for the completion of other writing projects.

In accepting this award Professor Gross thanked R. P. Boas, Editor of this MONTHLY, for his careful reading of the manuscript and thoughtful comments that resulted in a number of substantive expositional improvements. He also offered the following personal reflections upon those individuals who most significantly influenced his career. "My teacher, now colleague and friend, Ray A. Kunze, first introduced me to the subject of harmonic analysis. Through his patience and clarity of mathematical expression I was inspired to become a serious research mathematician. From my brother, Herbert I. Gross, a gifted mathematics educator and philosopher, currently at Bunker Hill Community College and MIT, I learned the importance of human values, as well as a sense of humor, in the art of teaching. By his model I have tried to elevate

the transmission of information, characteristic of our profession, to higher forms of education and human enrichment. In my friend and colleague Ernst Snapper, to whom this Chauvenet paper was dedicated, the aspiring young mathematician could see the rich and varied academic career available for those who apply high standards of excellence in all their professional activities. He is a fountain of eternal mathematical youth. Finally, while in graduate school I had the privilege of a close personal relationship with a professor, Guido Weiss, well known as an outstanding mathematical expositor. It is a pleasure to follow him, years later, in receiving the Chauvenet Prize. To these people, and many others unnamed, my debt is great."

David P. Roselle, *Secretary*

THE GEOMETRY OF ROOT SYSTEMS AND SIGNED GRAPHS

THOMAS ZASLAVSKY

Department of Mathematics, The Ohio State University, Columbus, OH 43210

Dedicated to Professor Fred Supnick
of The City College and the City University of New York:
A small return for much given.

This essay tells of a newly discovered connection among root systems, graphs, and matroids.

Root systems are sets of vectors which satisfy certain requirements of symmetry and metric regularity. They arose in Lie theory, where they are important because they correspond one-to-one to Lie algebras and hence to Lie groups and because many properties of the algebra and group involve the root system. They have since found other applications, such as to line graphs and the search for finite simple groups.[1]

Graphs, or networks, which consist of nodes joined by arcs, arise in all kinds of combinatorial analysis. Yet signed graphs, in which each arc is labeled by $+$ or $-$, are rarely discussed or applied. (So much so that some people at first think they are another form of directed graph. They are not.) They will find good use in this article, for with five exceptions every root system can be concisely and faithfully represented by a signed graph.

One of the problems encountered in Lie theory is that of counting the pieces into which space is cut by all the hyperplanes dual to elements of a root system. The usual method of solution, which is classical and well known, depends on translating the problem into one concerning an automorphism group of the root system. But it is not necessary to take that approach. Instead, by tackling the problem directly with combinatorial techniques, one can count the pieces derived not only from full root systems but from many subsystems; roughly speaking, the root systems correspond to complete graphs, while the additional systems solvable by combinatorics correspond to arbitrary subgraphs. The principal tool is the characteristic polynomial of an arrangement of hyperplanes (see Section 4), a polynomial borrowed from matroid theory. This theory, which I shall not need to mention again by name, is nonetheless the quiet ground of my discourse.

1. Root Systems. A *root system* is a finite set R of vectors in $\mathbb{R}^n$, called *roots*, with the properties:

The author received his Ph.D. from MIT under the direction of Curtis Greene. He has taught at MIT and is now at Ohio State University. His principal research interest is in combinatorial geometry of many kinds.–*Editors*

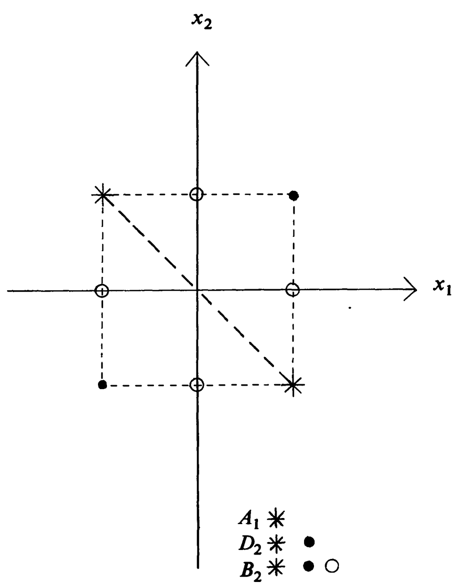
(1) if $\alpha \in R$, the only integral multiples $k\alpha$ which are in R are α and $-\alpha$ (thus 0 is not a root);

(2) if α, β are linearly independent roots and if $p, q \geqslant 0$ are the largest integers such that $\alpha + k\beta$ is a root for all integers k in the range $-p \leqslant k \leqslant q$, then $(\alpha,\beta)/(\alpha,\alpha) = -(p+q)$.

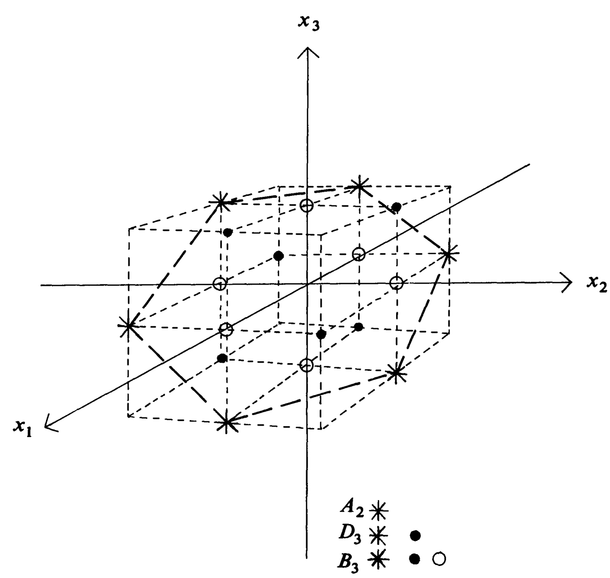
It turns out that there are very few root systems which cannot be built out of others in a simple way. Let us suppose $S \subseteq \mathbb{R}^l$ and $T \subseteq \mathbb{R}^m$ are root systems. Then their union $S \cup T$ is in a natural way a subset of $\mathbb{R}^{l+m}$, with S and T lying in orthogonal subspaces. This set $S \cup T$, the *direct sum* of S and T, is obviously also a root system. A root system which cannot be decomposed as a direct sum is *irreducible*. Now let's look at irreducible root systems. Two of them are called *similar*, and considered to be essentially identical, if there is a change of scale (a similarity transformation) which makes them isometrically isomorphic. The remarkable fact is that, aside from five "exceptional" root systems (known as G_2, F_4, E_6, E_7, and E_8), there are only four families of irreducible root systems. They are the *classical root systems* A_{n-1}, B_n, C_n, and D_n, which are traditionally represented in $\mathbb{R}^n$ as the vector sets

$$\begin{aligned} A_{n-1} &= \{b_i - b_j\}_{i\neq j},\\ D_n &= A_{n-1} \cup \{\pm(b_i + b_j)\}_{i \neq j},\\ B_n &= D_n \cup \{\pm b_i\},\\ C_n &= D_n \cup \{\pm 2b_i\}, \end{aligned}$$

where $b_1, b_2, \ldots, b_n$ are an orthonormal basis of $\mathbb{R}^n$. The proof of this classification theorem, which is long and complicated, appears in most books on Lie algebras. (The dimension of the system is indicated by its subscript; so all span $\mathbb{R}^n$ except A_{n-1}. In dimensions $\leqslant 3$ there are some similarity relations and reducibilities among the classical systems, which will not concern us.)



(a). The three root systems A_1, D_2, B_2 in $\mathbb{R}^2$.



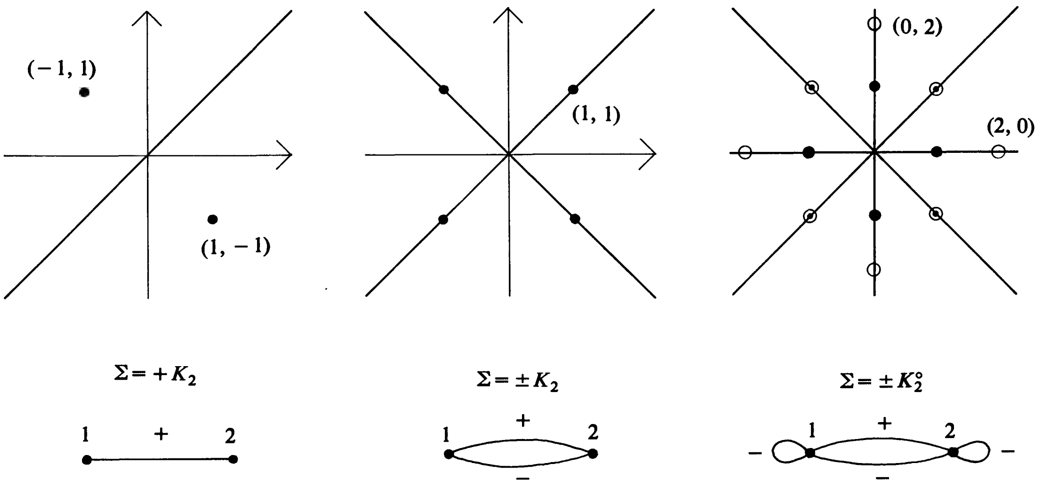
(b). The three root systems A_2, D_3, B_3 in $\mathbb{R}^3$.

FIG. 1

2. Arrangements, Regions, and Chambers. Let R be a root system in $\mathbb{R}^n$. If we take the

hyperplane perpendicular to each root we get a finite set of hyperplanes, R^*, which dissects $\mathbb{R}^n$ into n-dimensional pieces called *Weyl chambers*. They are the components of $\mathbb{R}^n \setminus \cup \{h \in R^*\}$.

We could easily be more general. Let H be any finite set of hyperplanes. We call H an *arrangement of hyperplanes* and the components of $\mathbb{R}^n \setminus \cup \{h \in H\}$ the *regions* of the arrangement. (If $H = R^*$ the "regions" are also "chambers"; here we see the meeting of two independent traditions.)



(a) A_1, A_1^*, and $\Sigma = +K_2$. There are 1 hyperplane (a line) and 2 chambers.

(b) D_2, D_2^*, and $\Sigma = \pm K_2$. There are 2 lines and 4 chambers.

(c) $B_2(\bullet)$, $C_2(\circ)$, $B_2^* = C_2^*$, and $\Sigma = \pm K_2^\circ$. There are 4 lines and 8 chambers.

FIG. 2. The classical root systems in $\mathbb{R}^2$, their dual arrangements of hyperplanes, and their corresponding signed graphs.

Let $c(H)$ denote the number of regions. In Lie theory the number of chambers, $c(R^*)$, is important enough to be the subject of a chapter. It is calculated by means of a certain symmetry group of R, the Weyl group, whose order equals the number of chambers. But it seemed to me that it should not be necessary to rely on the symmetry of R to count its chambers. Such a combinatorial problem should have a combinatorial solution. As it happened, I knew a purely combinatorial technique for counting the regions of an arrangement of hyperplanes without any assumption of symmetry. In the rest of this paper (after a brief classical interlude) I will show how to use it, by way of the medium of signed graphs, to find the numbers of chambers of all the classical root system arrangements as well as the numbers of regions of many of their subarrangements (not all, because the calculations become too complicated).

3. The Classical Approach to Weyl Chambers. In Lie theory the number of chambers of R, $c(R^*)$, is calculated by showing that a certain group $\mathfrak{W}(R)$, the *Weyl group*, permutes the chambers and that for each two chambers, C_1 and C_2, exactly one $w \in \mathfrak{W}(R)$ carries C_1 to C_2. Thus $c(R^*) =$ the order of $\mathfrak{W}(R)$. The Weyl group is generated by the reflections S_α for $\alpha \in R$, where S_α means orthogonal reflection of $\mathbb{R}^n$ in the hyperplane h_α perpendicular to α. It is easy to compute that

$$\mathfrak{W}(A_{n-1}) \cong \mathfrak{S}_n,$$
$$\mathfrak{W}(B_n) \cong \mathfrak{W}(C_n) \cong \mathfrak{O}_n,$$
$$\mathfrak{W}(D_n) \cong \mathfrak{O}_n^+,$$

where $\mathfrak{S}_n =$ the symmetric group on n letters, or the group of $n \times n$ permutation matrices; $\mathfrak{O}_n$ is

the hyperoctahedral group, or the group of $n\times n$ signed permutation matrices; and $\mathfrak{O}_n^+$ denotes the subgroup of $\mathfrak{O}_n$ consisting of the matrices with evenly many minus signs. From the one-to-one correspondence between elements of the Weyl group and the Weyl chambers, we deduce:

THEOREM 1. *$c(A_{n-1}^*)=n!$, $c(D_n^*)=2^{n-1}n!$, and $c(B_n^*)=c(C_n^*)=2^n n!$.*

Now let me show you how to calculate these same numbers without reference to the Weyl group.

4. The Combinatorial Approach to Regions. Here is the formula for the number of regions of an arrangement of hyperplanes in $\mathbb{R}^n$. Let H be the arrangement. Let us write $d(S)=\dim(\cap S)$ for any $S\subseteq H$. The *characteristic polynomial* of H is the polynomial defined by [2]

$$p_H(\lambda)=\sum_{S\subseteq H}(-1)^{\#(S)}\lambda^{d(S)-d(H)}.$$

THEOREM 2. *$c(H)=(-1)^{n-d(H)}p_H(-1)$.*

The proof of Theorem 2 is a good illustration of the powerful inductive method of deletion and contraction, which I will use again in the course of the paper.[3]

Pick a hyperplane $h\in H$. The other hyperplanes in H dissect h into $(n-1)$-dimensional pieces, which are the regions of the *induced arrangement* of hyperplanes in h,

$$H/h=\{h_1\cap h: h_1\in H, h_1\neq h\}.$$

H/h is the "contraction" of H to h. The "deletion" is $H\setminus h$. The method of deletion and contraction begins by proving two parallel equations:

$$c(H)=c(H\setminus h)+c(H/h) \tag{1}$$

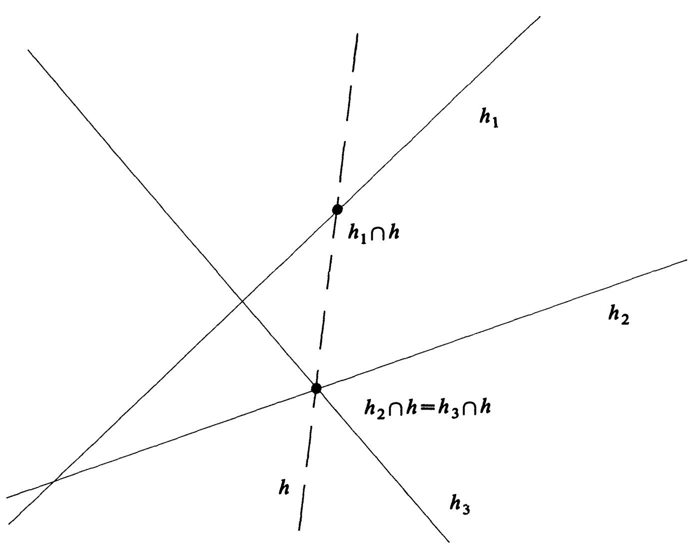


FIG. 3 (a). Removing a line h from an arrangement of lines $H=\{h,h_1,h_2,h_3\}$, yielding $H_1=\{h_1,h_2,h_3\}$. The induced arrangement H/h consists of two points on h. (In this example, because all hyperplanes do not pass through a common point, the definition of $p_H(\lambda)$ must be modified: see [14, pp. 3 and 13]. But the method of deletion and contraction remains valid.)

$$p_{H_1}(\lambda)=\lambda^2-3\lambda+3, \qquad c(H_1)=7=|p_{H_1}(-1)|,$$
$$p_{H/h}(\lambda)=\lambda-2, \qquad c(H/h)=3=|p_{H/h}(-1)|,$$
$$p_H(\lambda)=\lambda^2-4\lambda+5, \qquad c(H)=10=|p_H(-1)|.$$

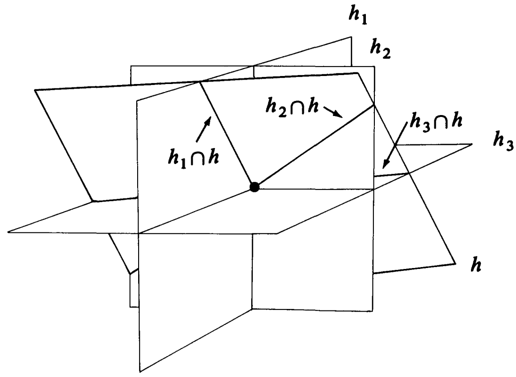


FIG. 3 (b). Removing a (hyper)plane h from an arrangement $H=\{h, h_1, h_2, h_3\}$ of planes, leaving $H_1=\{h_1, h_2, h_3\}$. The induced arrangement H/h consists of 3 lines in h, all passing through a common point.

$$p_{H_1}(\lambda)=\lambda^3-3\lambda^2+3\lambda-1, \qquad c(H_1)=8,$$
$$p_{H/h}(\lambda)=\lambda^2-3\lambda+2, \qquad c(H/h)=6,$$
$$p_H(\lambda)=\lambda^3-4\lambda^2+6\lambda-3, \qquad c(H)=14.$$

for any $h \in H$, and

$$p_H(\lambda)=p_{H\setminus h}(\lambda)-p_{H/h}(\lambda) \tag{2}$$

for any $h \in H$ such that $\cap H = \cap(H\setminus h)$, or $h \supseteq \cap(H\setminus h)$. Since the deletion and the contraction have fewer hyperplanes than H, we can carry out induction on the size of H, deducing Theorem 2 for H from its validity for $H\setminus h$ and H/h. The only hitch will be when (2) does not apply, which is the case when no $h \in H$ contains $\cap(H\setminus h)$. But if that is so, H must consist of $n-d(H)$ hyperplanes in general position; it is easy to see that $c(H)=2^{\#(H)}$ and it is an easy calculation that $p_H(\lambda)=(\lambda-1)^{\#(H)}$ (since $d(S)=n-\#(S)$). Now clearly Theorem 2 is valid.

The heart of the proof is thus the verification of Equations (1) and (2). You will have no trouble carrying out the rather long calculations necessary for (2) once I point out that there is a one-to-one correspondence between subarrangements $S \subseteq H$ which contain h and subarrangements $S' \subseteq H/h$, namely, $S \leftrightarrow S/h$, and that $\#(S)=\#(S')+1$ and $\cap S = \cap S'$. (It is only fair to admit there are some complications. There may be several hyperplanes which coincide; nevertheless they have to be treated as distinct objects. And one must allow the degenerate "hyperplane," which is the whole space—an arrangement containing it has no regions.)

By contrast Equation (1) is purely pictorial. Each region of H arises from a region C of $H\setminus h$. If h hits C, it cuts it into two parts, C_1 and C_2, which are regions of H, and one $(n-1)$-dimensional piece, $h \cap C$, which is a region of H/h. All the regions of H/h come about this way. If h misses C, however, then C is a region of H. Adding everything up, we get (1). That proves everything we needed for Theorem 2.

5. Equations and Signed Graphs. How does all this apply to the root system arrangements? We have to have some way of calculating their characteristic polynomials. With the help of signed graphs we'll be able to compute $p_{A^*_{n-1}}$, $p_{D^*_n}$, $p_{B^*_n}=p_{C^*_n}$, and more besides.[4]

Let's look at the hyperplanes of B^*_n. Each of them has one of the two forms

$$h^{\varepsilon}_{ij}\colon x_i=\varepsilon x_j \qquad \text{where } \varepsilon=\pm 1 \text{ and } i \neq j,$$
$$h_i\colon x_i=0.$$

(I think of a hyperplane as being the same as its equation for all practical purposes.) There is a nice, compact way of describing this. Let us construct a signed graph Σ on the n nodes

$\{1,2,\ldots,n\}$. Each node i corresponds to the coordinate x_i. A hyperplane (or equation) h_{ij}^{ε} of the first type corresponds to an arc of Σ which links i to j. To distinguish h_{ij}^{+} from h_{ij}^{-} we label the arc by the sign of h_{ij}^{ε}. Call this arc e_{ij}^{ε}. As for a hyperplane h_i of the second type, it is the same as $h_{ii}^{-}: x_i = -x_i$ so it corresponds to an arc e_{ii}^{-} (a negative loop at i). We have the following table:

$$h_{ij}^{\varepsilon}: x_i = \varepsilon x_j \leftrightarrow e_{ij}^{\varepsilon}: \text{linking } i \text{ and } j,$$

$$h_i: x_i = 0 \leftrightarrow e_{ii}^{-}: \text{loop at } i.$$

Note that according to our scheme a positive loop e_{ii}^{+} corresponds to the equation $x_i = x_i$, which is the whole space (the "degenerate hyperplane"). A negative loop, however, is a real hyperplane.

So if we take any subarrangement $H \subseteq B_n^*$ we can describe it by a signed graph. Conversely any signed graph Σ on the n nodes $\{1,2,\ldots,n\}$ describes a unique arrangement,

$$H[\Sigma] = \{h_{ij}^{\varepsilon}: \text{there is an arc } e_{ij}^{\varepsilon} \text{ in } \Sigma\}.$$

What we wanted from all this was a way to calculate $p_H(\lambda)$. It can be written down directly from the signed graph Σ corresponding to H, but as doing so for arbitrary H requires some relatively esoteric concepts, I will skip the general solution.[5] Instead I will discuss two kinds of subarrangement for which the value $p_H(\lambda)$ is particularly intriguing and easy to formulate.

6. "Special" Subarrangements. Let's call a subarrangement $H \subseteq B_n^*$ "special" if it contains all the coordinate hyperplanes and it has *sign symmetry*: whenever $h_{ij}^{\varepsilon} \in H$ (for $i \neq j$), also $h_{ij}^{-\varepsilon} \in H$.

The "special" arrangements correspond to the signed graphs I call "full signed expansions of ordinary graphs." Let Γ be an ordinary graph (for the sake of simplicity, without loops or multiple arcs). By $\pm\Gamma$ I mean the signed graph which has the same nodes as Γ and, for each arc e_{ij} of Γ, both the signed arcs e_{ij}^{+} and e_{ij}^{-}. This graph is *sign-symmetric* because it contains $e_{ij}^{-\varepsilon}$ whenever it contains e_{ij}^{ε} (for $i \neq j$). By $(\pm\Gamma)^\circ$, loosely written just $\pm\Gamma^\circ$, I mean the signed graph $\pm\Gamma$ with a negative loop added to every node. Having all these loops makes $\pm\Gamma^\circ$ a *full* sign-symmetric signed graph. The "special" arrangements are just those which equal $H[\pm\Gamma^\circ]$ for some ordinary graph Γ. For example, $B_n^* = H[\pm K_n^\circ]$. The point of singling them out is the simplicity of their characteristic polynomials. To state the theorem we need the *chromatic polynomial* $\chi_\Gamma(\lambda)$ of the graph Γ. If λ is a positive integer, $\chi_\Gamma(\lambda)$ is the number of proper colorings of Γ by λ colors: which means assigning an integer from the set $\{1,2,\ldots,\lambda\}$ to each node so that the two endpoints of an arc have different colors. The function χ_Γ turns out to be a polynomial; this lets us define it for other values of λ besides positive integers.

THEOREM 3. *Let Γ be a graph on the nodes $\{1,2,\ldots,n\}$ and $H = H[\pm\Gamma^\circ]$. Then*

$$p_H(\lambda) = 2^n \chi_\Gamma\left(\frac{\lambda-1}{2}\right). \tag{3}$$

In particular

$$p_{B_n^*}(\lambda) = p_{C_n^*}(\lambda) = 2^n \left(\frac{\lambda-1}{2}\right)_n,$$

where $(\lambda)_n$ denotes the falling factorial $\lambda(\lambda-1)\cdots(\lambda-n+1)$.

The proof of the particular case depends on the observation that $\chi_{K_n}(\lambda) = (\lambda)_n$.

I will prove the theorem by again applying deletion and contraction. The necessary recursion for the right-hand side is both well known and easy to prove. It is

$$\chi_\Gamma(\lambda) = \chi_{\Gamma\setminus e}(\lambda) - \chi_{\Gamma/e}(\lambda) \tag{4}$$

for $e = e_{ij}$ = any arc of Γ (except a loop, which we've ruled out anyway), where $\Gamma\setminus e$ is Γ with e removed and Γ/e means Γ "contracted" by e: the endpoints of e are merged and e itself is thrown away. We can prove (4) for every positive integer λ by counting colorings of Γ by λ

colors. If we ignore $e=e_{ij}$, letting i and j be colored the same, we have $\chi_{\Gamma\backslash e}(\lambda)$ colorings. Some of them are proper for Γ (i and j are colored differently); there are $\chi_{\Gamma}(\lambda)$ of these. The rest are the ways to color $\Gamma\backslash e$ so that i and j have the same color; their number equals $\chi_{\Gamma/e}(\lambda)$. Thus (4) holds for all positive integers λ. Since $\chi_{\Gamma}(\lambda)$ is a polynomial, it follows that (4) is a polynomial identity, valid for all λ.[6]

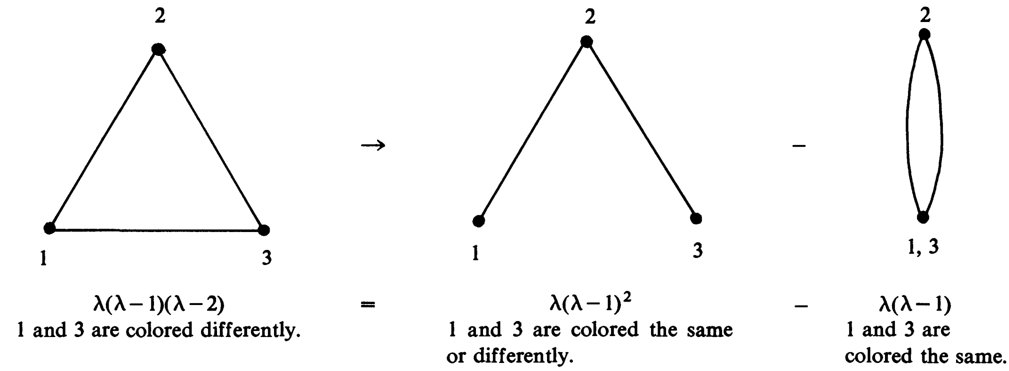


FIG. 4. Graph coloring, showing deletion and contraction. Under each graph is its chromatic polynomial.

Now we can prove Theorem 3 by induction on the number of arcs in Γ. We begin, of course, with the arcless graph on the node set $\{1,2,\ldots,n\}$; call it V_n. Its chromatic polynomial is $\chi_{V_n}(\lambda)=\lambda^n$. The corresponding "special" arrangement $H[\pm V_n^\circ]$ consists of the coordinate hyperplanes. Its characteristic polynomial is easily seen from the definition to be $(\lambda-1)^n$. This equals $2^n\chi_{V_n}(\frac{1}{2}(\lambda-1))$, as we wanted.

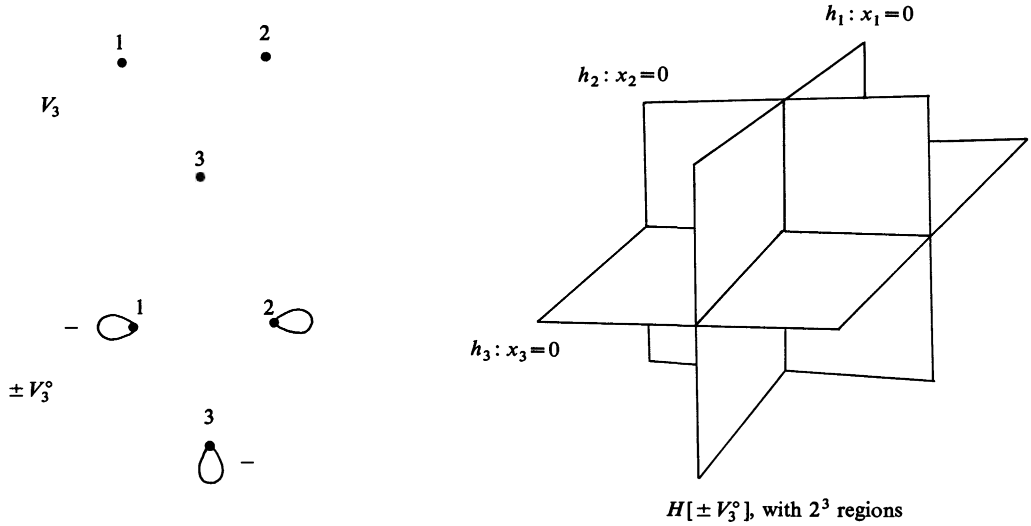
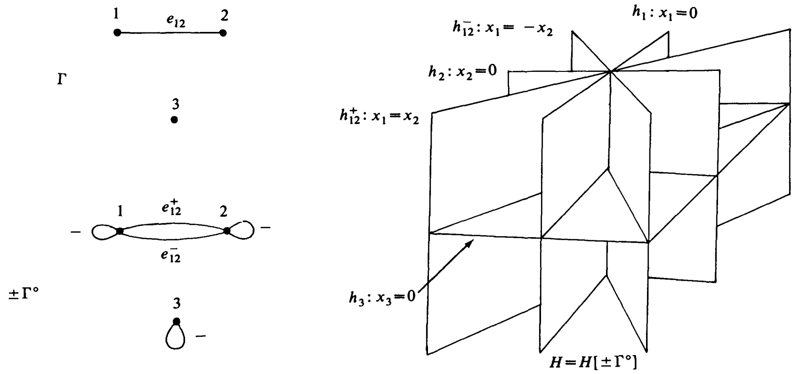


FIG. 5. The graph V_3, the signed graph $\pm V_3^\circ$, and the "special" arrangement $H[\pm V_3^\circ]$.

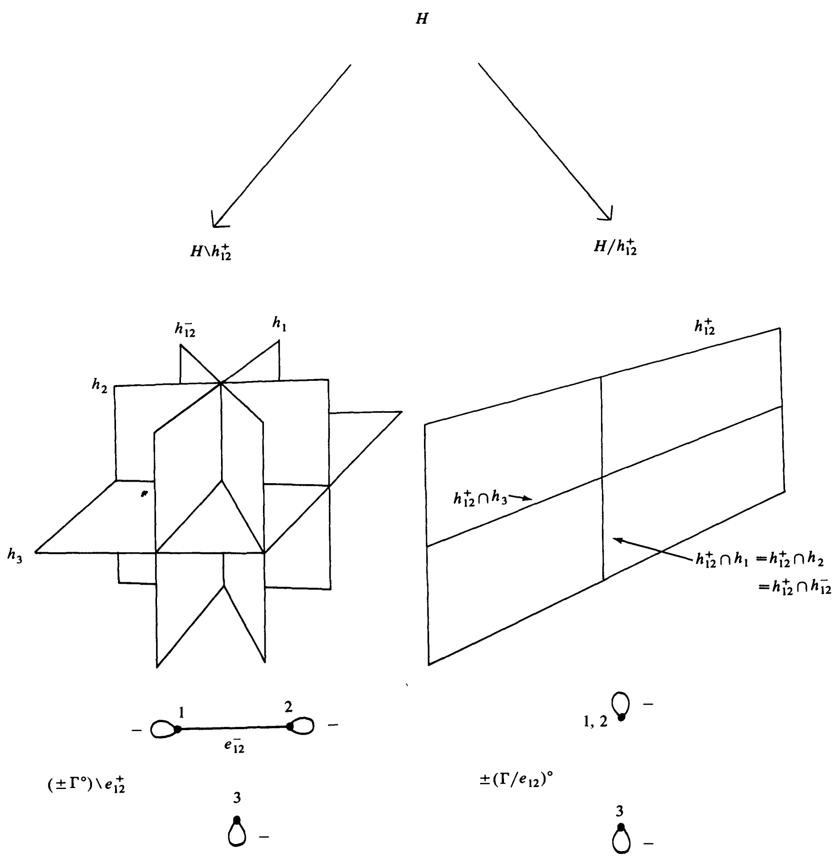
Next we carry out the induction. It is a relatively complicated application of deletion and contraction: it uses it twice. First we pick one arc e_{ij} of Γ and apply deletion and contraction by h_{ij}^+ to $H=H[\pm\Gamma^\circ]$, using Equation (2). The result is

$$p_H(\lambda)=p(H\backslash h_{ij}^+;\lambda)-p(H/h_{ij}^+;\lambda). \tag{5}$$

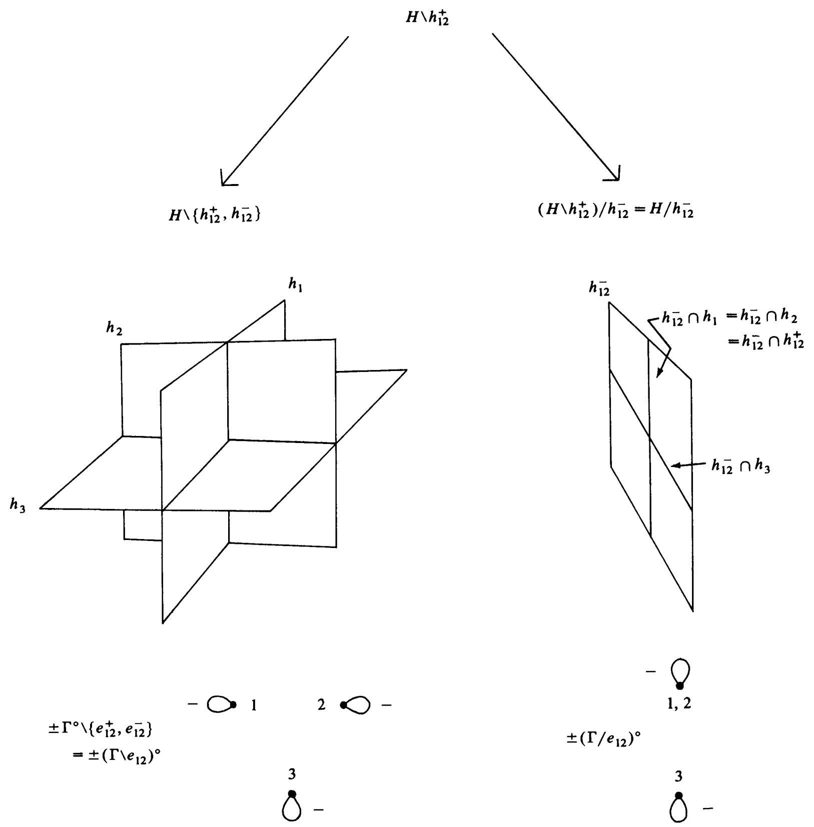
FIG. 6. An example of the inductive procedure used to prove Theorem 3.



(a) The initial graph Γ, full signed expansion graph $\pm\Gamma^\circ$, and "special" arrangement $H[\pm\Gamma^\circ]$.



(b) The arrangements resulting from deletion of and contraction by h_{12}^{+}, together with the signed graphs which describe them.



(c) The arrangements resulting from deletion and contraction of $H\backslash h_{12}^{+}$ by h_{12}^{-}, together with the signed graphs which describe them.

Then we do the same to $H\backslash h_{ij}^{+}$, deleting and contracting with respect to h_{ij}^{-}. We obtain

$$p(H\backslash h_{ij}^{+};\lambda)=p(H\backslash\{h_{ij}^{+},h_{ij}^{-}\};\lambda)-p((H\backslash h_{ij}^{+})/h_{ij}^{-};\lambda). \tag{6}$$

Obviously $H\backslash\{h_{ij}^{+},h_{ij}^{-}\}=H[\pm(\Gamma\backslash e_{ij})^{\circ}]$, which is a smaller "special" arrangement than H. But what about the contractions H/h_{ij}^{+} and $(H\backslash h_{ij}^{+})/h_{ij}^{-}$?

They are also "special"; they both equal $H[\pm(\Gamma/e_{ij})^{\circ}]$. Let us see why. The arrangement H/h_{ij}^{+} is the one induced by H on $h_{ij}^{+}: x_i=x_j$. That means we take every equation in H and set x_j identically equal to x_i. An equation $x_k=\varepsilon x_j$ becomes the same as $x_k=\varepsilon x_i$; $h_j: x_j=0$ and $h_{ij}^{-}: x_i=-x_j$ both become repetitions of $h_i: x_i=0$. The effect is as if we had merged i and j in Γ. (Notice that h_{ij}^{-} becomes the same as h_i, so that $(H\backslash h_{ij}^{-})/h_{ij}^{+}=H/h_{ij}^{+}$. By the sign symmetry of the situation, $(H\backslash h_{ij}^{+})/h_{ij}^{-}=H/h_{ij}^{+}$.) Thus we have proved

$$H/h_{ij}^{+}=(H\backslash h_{ij}^{+})/h_{ij}^{-}=H[\pm(\Gamma/e)^{\circ}].$$

In the light of this fact, combining (5) and (6) we have

$$p_{H[\pm\Gamma^{\circ}]}(\lambda)=p_{H[\pm(\Gamma\backslash e)^{\circ}]}(\lambda)-2p_{H[\pm(\Gamma/e)^{\circ}]}(\lambda). \tag{7}$$

Equation (7) is what we need to do induction, since $\Gamma\backslash e$ and Γ/e have fewer arcs than Γ. We can complete the proof by substituting from (3) in the right-hand side of (7)—remembering that Γ/e has $n-1$ nodes.

COROLLARY 4. *$H[\pm\Gamma°]$ has $2^n|\chi_\Gamma(-1)|$ regions. In particular $c(B_n^*)=2^n n!$.*

The proof is by Theorems 2 and 3.

Notice how we have gotten $c(B_n^*)$ without any group theory. As a bonus we see that the 2^n and the $n!$ enter into $c(B_n^*)$ for different reasons: the 2^n because B_n^* is "special," the $n!$ because it derives from the complete graph.

7. Graphic Subarrangements. A subarrangement of B_n^* is called *graphic* if it contains only hyperplanes with plus signs, $h_{ij}^+ : x_i = x_j$. If Γ is a graph with node set $\{1,2,\dots,n\}$, let

$$H[\Gamma]=\{h_{ij}^+ : \text{there is an arc } e_{ij} \text{ in } \Gamma\}.$$

Obviously $H[\Gamma]$ is graphic; conversely, if H is a graphic arrangement, it is derived from the graph whose arcs are $\{e_{ij}: h_{ij}^+ \in H\}$. This is the reason for the name "graphic."

THEOREM 5. *Let Γ be a graph on the nodes $\{1,2,\dots,n\}$, having c components, and let $\chi_\Gamma(\lambda)$ be its chromatic polynomial. Then*

$$p_{H[\Gamma]}(\lambda)=\chi_\Gamma(\lambda)/\lambda^c.$$

In particular

$$p_{A_{n-1}^*}(\lambda)=(\lambda-1)_{n-1}=(\lambda-1)(\lambda-2)\cdots(\lambda-n+1).$$

COROLLARY 6. *$H[\Gamma]$ has $|\chi_\Gamma(-1)|$ regions. In particular $c(A_{n-1}^*)=n!$.*

The theorem is a consequence of the fact that $H[\Gamma]$ represents the graphic geometry (or polygon matroid) of Γ.[7]

Corollary 6 was first noticed by Curtis Greene. He also saw that the corollary can be strengthened: there is a one-to-one correspondence between the regions of $H[\Gamma]$ and the "acyclic orientations" of Γ. (It was these discoveries of Greene's, dating from 1975, that interested me in graphic arrangements.) An *acyclic orientation* of Γ is a way of directing the arcs so there is no closed path which follows their directions. There is an analogous, although more complex, interpretation of the regions of $H[\pm\Gamma°]$ in terms of Γ. There is also a close connection between the graphic and "special" arrangements associated with Γ. The fact that

$$c(H[\pm\Gamma°])=2^n c(H[\Gamma])$$

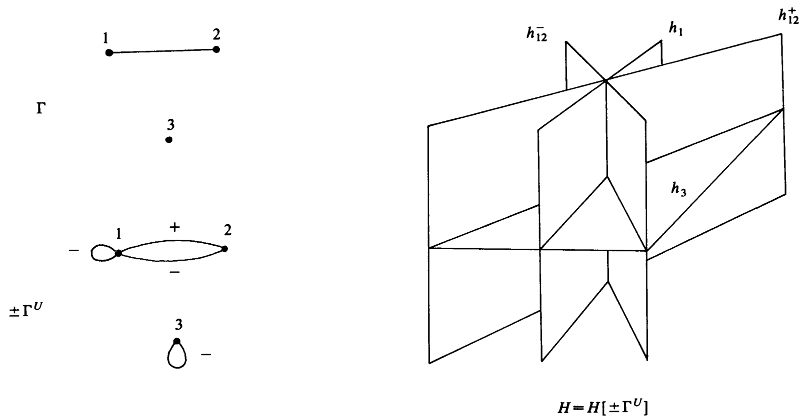
suggests that the regions of $H[\Gamma]$ may be individually sliced into 2^n parts each by the extra hyperplanes in $H[\pm\Gamma°]$. And so they are—although not, as one might think, through successive halving of each old region by n of the new hyperplanes.

8. Subarrangements with Sign Symmetry. So far we've counted the chambers of B_n^* and A_{n-1}^* as special cases of the "special" and graphic arrangements, $H[\pm\Gamma°]$ and $H[\Gamma]$. Although D_n^* belongs to neither of these types, it too is highly symmetrical: it has sign symmetry, for $D_n^* = H[\pm K_n]$. I will now show how to calculate the number of regions of a general sign-symmetric arrangement; in particular, that will take care of D_n^* and all $H[\pm\Gamma]$.

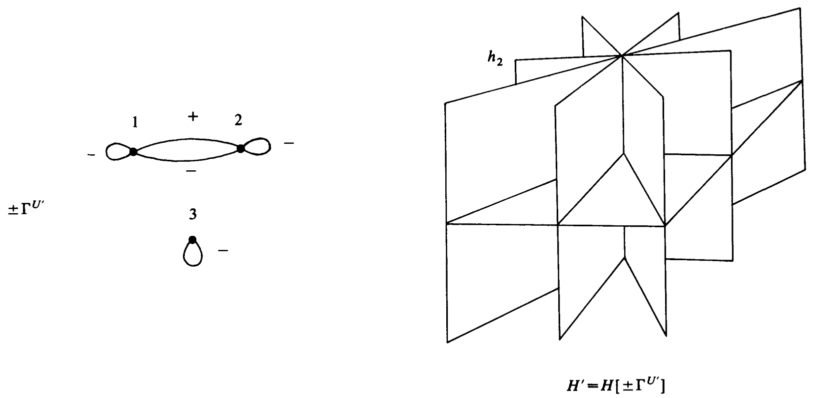
A sign-symmetric arrangement $H\subseteq B_n^*$ has associated to it an ordinary graph Γ, defined as having an arc e_{ij} linking distinct nodes i and j whenever H has hyperplanes h_{ij}^+ and h_{ij}^-. Suppose H corresponds to the signed graph Σ. Then Σ contains $\pm\Gamma$ and, since H consists of $H[\pm\Gamma]$ plus perhaps some coordinate hyperplanes, the remaining arcs of Σ must all be negative loops. We will need a way to describe Σ. So if U is a subset of the node set, let $\pm\Gamma^U$ be the signed graph which has all the arcs of $\pm\Gamma$ and in addition a negative loop at each node in U. Now we can say: Every sign-symmetric subarrangement of B_n^* is $H[\pm\Gamma^U]$ for some Γ and U. Conversely, every $H[\pm\Gamma^U]$ is obviously sign-symmetric.

What we want is a computational scheme for all arrangements of this form. That means we need to find the characteristic polynomials of all $H[\pm\Gamma^U]$. The problem can be reduced to the

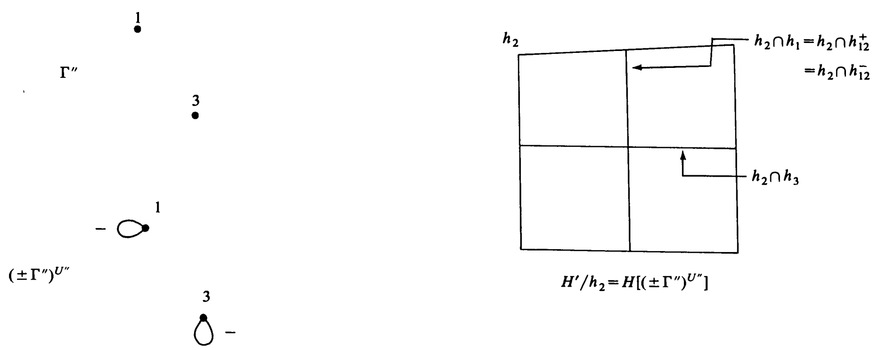
FIG. 7. An illustration of Theorem 7 in the case $q=0$ (all isolated nodes of Γ are in U).



(a) The original arrangement $H=H[\pm\Gamma^U]$, where $U=\{1,3\}$.



(b) The enlarged arrangement $H'=H\cup\{h_2\}$, which equals $H[\pm\Gamma^{U'}]$ where $U'=\{1,2,3\}$. Here $i=2$ and the added hyperplane is h_2.



(c) The contracted arrangement H'/h_2, which equals $H[(\pm\Gamma'')^{U''}]$ where $\Gamma''=\Gamma:\{2\}^c$ and $U''=\{1, 3\}$.

"special" case. If W is a set of nodes of Γ, the *subgraph induced by* W, written $\Gamma: W$, is the graph whose node set is W and whose arcs are all those of Γ both of whose endpoints are in W. In case there are no such arcs, we call W a *stable* set of nodes.

A node is *isolated* in Γ if it lies on no arcs.

THEOREM 7. *Let Γ be a graph on the nodes $\{1,2,\ldots,n\}$, U be any subset of the node set, $q=$ the number of isolated nodes of Γ which lie outside U, and $H=H[\pm\Gamma^U]$. Then*

$$p_H(\lambda)=\lambda^{-q}\sum_W 2^{\#(W)}\chi_{\Gamma:W}\left(\frac{\lambda-1}{2}\right), \tag{8}$$

where the sum is taken over every node set $W\supseteq U$ whose complement is stable in Γ. In particular

$$p_{D_n^*}(\lambda)=2^n\left(\frac{\lambda-1}{2}\right)_n+n2^{n-1}\left(\frac{\lambda-1}{2}\right)_{n-1}.$$

Let me write, informally, $H: W$ for the arrangement $H[\pm(\Gamma: W)^\circ]$, which lies in $\mathbb{R}^W$. If you refer to Theorem 3 you will see that Theorem 7 is equivalent to the formula

$$p_H(\lambda)=\lambda^{-q}\sum_W p_{H:W}(\lambda). \tag{9}$$

I will prove (8) in two stages: via (9) for the case $q=0$ by deletion and contraction (again!) along with induction on the number of nodes not in U (which is the number of coordinate hyperplanes not in H); then (8) directly for $q>0$ by an *ad hoc* reduction to the first case.

To begin with, if all nodes are in U, the range of W in the summation is merely $W=\{1,2,\ldots,n\}$. So (9) is trivially true.

If $q=0$ but U is not everything, pick a fixed $i\notin U$; write $U'=U\cup\{i\}$ and $H'=H[\pm\Gamma^{U'}]$. By deletion and contraction,

$$p_H(\lambda)=p_{H'}(\lambda)+p_{H'/h_i}(\lambda). \tag{10}$$

(Our appeal to deletion and contraction depends on the fact that $\cap H'=\cap H$. This is true because, since $i\notin U$, i is not isolated; so there is an arc e_{ij}. Therefore h_{ij}^+ and h_{ij}^- are in H, so $h_i\supseteq h_{ij}^+\cap h_{ij}^-\supseteq\cap H$.)

Now we have to know what H'/h_i is. But it is merely the arrangement in h_i which results from setting $x_i=0$ in all the hyperplanes of H'. An equation h_j: $x_j=0$ or h_{jk}^ε: $x_j=\varepsilon x_k$ in which $j,k\neq i$ will remain the same, but h_{ij}^ε: $x_i=\varepsilon x_j$ becomes transformed to h_j: $x_j=0$. That is, for every node j which is adjacent to i in Γ, H'/h_i will contain the corresponding coordinate hyperplane. In the language of signed graphs: Set $\Gamma''=\Gamma:\{i\}^c$ (where $\{i\}^c$ means the complement of $\{i\}$) and $U''=U\cup\{j: j\neq i$ is adjacent to i in $\Gamma\}$. Then $H'/h_i=H[(\pm\Gamma'')^{U''}]$.

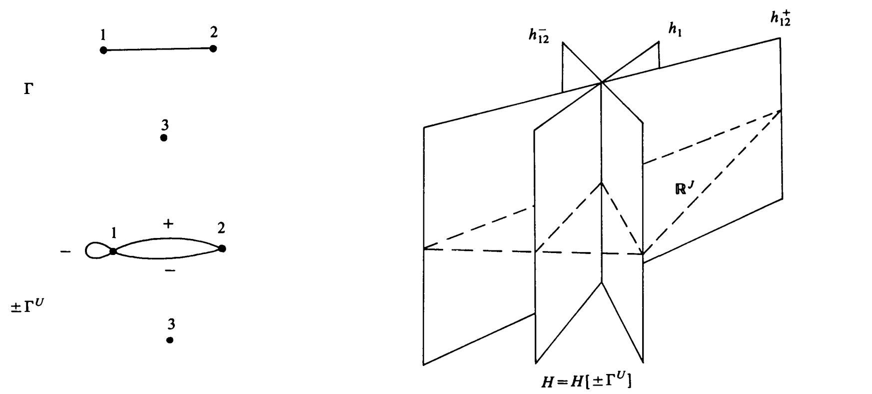
Notice that H' and H'/h_i have fewer missing coordinate hyperplanes than H. Also their signed graphs have no isolated nodes outside U' or U''. Hence we can use induction to assume the validity of (9) for H' and H'/h_i. Substituting in (10) leads to Equation (9) for H. Thus we have proved Theorem 7 for the case $q=0$.

What if $q>0$? Let Q be the set of isolated nodes of Γ which are not in U; and let J be its complement in the full node set, Q^c. I will reduce both sides of (8) to $\Gamma: J$ instead of Γ.

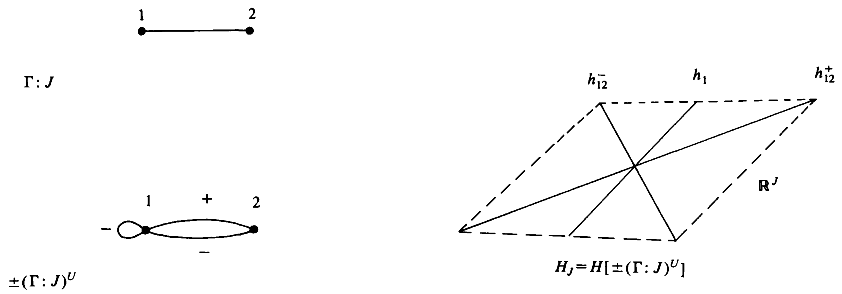
That means H will be replaced by $H_J=H[\pm(\Gamma: J)^U]$, the faithful cross section of H obtained by forgetting the Q coordinates. Everything in H_J is the same as in H except that all dimensions are lowered by q. Since that doesn't affect the characteristic polynomial, we have $p_H(\lambda)=p_{H_J}(\lambda)$.

On the other side of (8), let's split W into $Y=W\cap Q$ and $Z=W\cap J$. Since $\Gamma: W$ is the disjoint union of $\Gamma: Z$ and $\#(Y)$ isolated nodes, $\chi_{\Gamma:W}(\lambda)=\lambda^{\#(Y)}\chi_{\Gamma:Z}(\lambda)$. Moreover $W\supseteq U$ and W^c is stable in Γ, if and only if $Z\supseteq U$ and $J\setminus Z$ is stable in $\Gamma: J$. From this it is easy to check that

$$\sum_W 2^{\#(W)}\chi_{\Gamma:W}\left(\frac{\lambda-1}{2}\right)=\lambda^q\sum_{Z\subseteq J}2^{\#(Z)}\chi_{\Gamma:Z}\left(\frac{\lambda-1}{2}\right).$$



(a) The original arrangement $H=H[\pm\Gamma^U]$, where $U=\{1\}$. The set of isolated nodes not in U is $Q=\{3\}$; $q=1$; and $J=Q^c=\{1, 2\}$.



(b) The arrangement H_J, the cross section of H by $\mathbb{R}^J$. It is combinatorially equivalent to H.

FIG. 8. An illustration of Theorem 7 in the case $q>0$ (there are isolated nodes of Γ which are not in U).

That is the reduction we wanted to $\Gamma: J$. Thus the general part of Theorem 7 is proved.

Evaluating (8) for $D_n^* = H[\pm K_n]$, the stable node sets are $\varnothing$ and the singletons. So $W=\{1,2,\ldots,n\}$, giving $\Gamma: W=K_n$, and $W=\{i\}^c$ for $i=1,2,\ldots,n$, giving $\Gamma: W=K_{n-1}$. Since $q=0$, that gives us the particular formula.

COROLLARY 8. *Let Γ be a graph on the nodes $\{1,2,\ldots,n\}$, U be a subset of the node set, and $H=H[\pm\Gamma^U]$. Then H has*

$$c(H)=\sum_W (-1)^{n-\#(W)}2^{\#(W)}|\chi_{\Gamma: W}(-1)|$$

regions, where W ranges over every node set $W\supseteq U$ whose complement is stable in Γ. If $U=\varnothing$, W ranges over all complements of stable node sets. In particular $c(D_n^*)=2^n(n-1)!$.

The proof, of course, is by combining Theorems 2 and 7. As for D_n^*, we have

$$\begin{aligned} c(D_n^*) &= 2^n|(-1)_n| + n(-1)2^{n-1}|(-1)_{n-1}| \\ &= 2^n n! - n2^{n-1}(n-1)! \\ &= 2^{n-1}n!. \end{aligned}$$

This calculation is combinatorially interesting. It suggests that, rather than each chamber of B_n^* being the union of two from D_n^* as one might have thought from the formula $c(B_n^*)=2c(D_n^*)$, it is more likely that some chambers of D_n^* are not subdivided when one adds the coordinate hyperplanes while others are divided into three or more pieces. Analyzing the characteristic inequalities defining each chamber shows this to be true. One can also see that each successive

coordinate hyperplane added to D_n^* halves $2^{n-1}(n-1)!$ regions. (*Proof.* Let $D_n^{*(k)}$ be D_n^* with k coordinate hyperplanes added. Then

$$c(D_n^{*(k)}) = 2^{n-1}n! + k2^{n-1}(n-1)!$$

by Corollary 8. So in $D_n^{*(k+1)}$, $2^{n-1}(n-1)!$ regions of $D_n^{*(k)}$ are halved.) Is there any significance attached to which chambers of D_n^* are cut by any given number of coordinate hyperplanes or are separated into a given number of chambers of B_n^*?

9. Faces and Flats. With all this lengthy discussion of regions and chambers, I have not yet mentioned the lower-dimensional faces of an arrangement of hyperplanes. Take a region of an arrangement H in $\mathbb{R}^n$. It is an n-dimensional convex polyhedron, with flat sides, not bounded since it is a cone radiating from the origin—rather like an infinite wedge or pyramid. Each flat side (of whatever dimension) is called a *face* of the region and of the arrangement.

I can define faces more precisely by applying the notion of a *flat* of H: a subspace which is the intersection of hyperplanes in H. I include as flats the whole space and each hyperplane. One way to define a face is as the relative interior of an intersection $\bar{C} \cap t$, where $\bar{C}$ is the topological closure of a region C and t is any flat. Another way (completely equivalent) is as any region of any arrangement H/t induced by H on a flat. The largest faces are the regions. The smallest face is the intersection of all the hyperplanes; in most cases this is the origin (but not for graphic arrangements, where it always contains the line $x_1 = \cdots = x_n$).

For instance D_2^* (see Fig. 2(b)) has 4 regions, 4 one-dimensional faces (they are rays from the origin), and 1 zero-dimensional face (the origin).

Remarkably enough, with little extra effort we can calculate the number of k-dimensional faces for any k, which is denoted f_k, and the number of k-dimensional flats, denoted a_k, for all sign-symmetric root system subarrangements. The secret weapon is a certain polynomial, the *Whitney polynomial* of H,[8]

$$w_H(x,\lambda) = \sum_{T \subseteq S \subseteq H}\sum x^{n-d(T)}(-1)^{\#(S)-\#(T)}\lambda^{d(S)-d(H)},$$

where $d(S)$, you may recall, $= \dim(\cap S)$. Now f_k is equal to the coefficient of x^{n-k} in $(-1)^{n-d(H)}w_H(-x,-1)$; while the coefficient of $x^{n-k}\lambda^{k-d(H)}$ in $w_H(x,\lambda)$ is a_k. To see why, use a second version of w_H,

$$w_H(x,\lambda) = \sum_t x^{n-\dim t} p_{H/t}(\lambda), \tag{11}$$

summed over all flats of H; the second definition of a face, from which it follows that

$$f_k = \sum_{t:\dim t = k} c(H/t);$$

and Theorem 2 applied to H/t. (*Proof of Equation* (11): Fix T; let $t = \cap T$ and $T_1 = \{h \in H \setminus T : h \supseteq t\}$. If $T_1 \neq \varnothing$ then the sum over all $S \supseteq T$ will equal 0. On the other hand, the T such that $T_1 = \varnothing$ are in one-to-one correspondence with the flats of H.)

Our problem, then, is to compute the Whitney polynomial. So long as we stick to the arrangements we've investigated—the graphic and sign-symmetric subarrangements of the root systems—and remember Theorems 3, 5, and 7, the calculations are quite straightforward. I will skip them and just give their results.[9] The *contraction* Γ/T of a graph Γ by an arc set T is the graph resulting from coalescing each group of nodes which are connected by T and then discarding T; the arcs of Γ/T are thus $E(\Gamma)\setminus T$. The *Whitney polynomial of the graph* Γ is

$$w_\Gamma(x,\lambda) = \sum_{T \subseteq E(\Gamma)} x^{c(T)}\chi_{\Gamma/T}(\lambda),$$

$c(T)$ being the number of components into which T connects the nodes of Γ (it is the number of

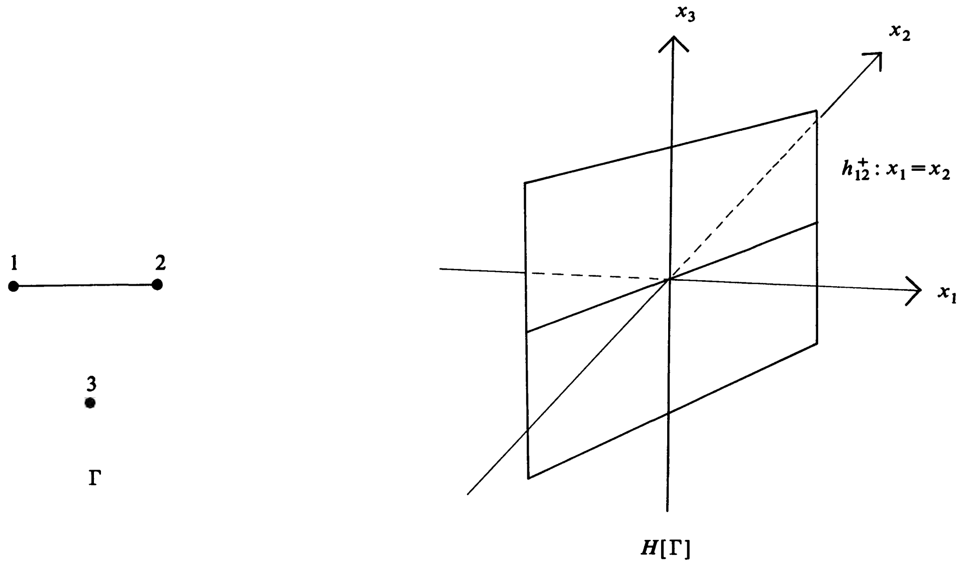


FIG. 9. A sample calculation of $\alpha_k(\Gamma)$ and $\phi_k(\Gamma)$ for use in Corollaries 10 and 11. You can see by inspecting Γ that

$\alpha_1(\Gamma)=0$ since Γ is not connected,
$\alpha_2(\Gamma)=1$ since N has one partition into 2 connected blocks,
$\alpha_3(\Gamma)=1$ since N has one partition into 3 connected blocks.

The ϕ_k can (says Corollary 11) be seen in the picture of $H[\Gamma]$:

$\phi_0(\Gamma)=0$ since $H[\Gamma]$ has no vertices,
$\phi_1(\Gamma)=0$ since it has no 1-faces (rays or edges),
$\phi_2(\Gamma)=1$ since it has one 2-face (the plane h_{12}^+),
$\phi_3(\Gamma)=2$ since h_{12}^+ divides the space into two regions.

nodes of Γ/T). I should explain that $\chi_{\Gamma/T}(\lambda)\equiv 0$ if there is an arc $e\notin T$ whose endpoints are connected by T; because Γ/T then has a loop, so no colorings. If there is no such arc, T is called *closed*. Then $\chi_{\Gamma/T}(\lambda)$ is monic of degree $c(T)$.

THEOREM 9. *Let Γ be a graph on the node set $N=\{1,2,\ldots,n\}$ and $c(\Gamma)=$ the number of components of Γ. The Whitney polynomial of the graphic arrangement $H[\Gamma]$ is*

$$w_{H[\Gamma]}(x,\lambda)=\lambda^{-c(\Gamma)}w_\Gamma(x,\lambda).$$

That of the "special" arrangement $H=H[\pm\Gamma^\circ]$ is

$$w_H(x,\lambda)=\sum_{W\subseteq N} x^{n-\#(W)}2^{\#(W)}w_{\Gamma:W}\left(x,\frac{\lambda-1}{2}\right).$$

Let U be a subset of the node set. The Whitney polynomial of the sign-symmetric arrangement $H=H[\pm\Gamma^U]$ is

$$w_H(x,\lambda)=\lambda^{-i(N)}\sum_{W\subseteq N} x^{\#(W^c)-i(W^c)}2^{\#(W)}w_{\Gamma:W}\left(x,\frac{\lambda-1}{2}\right),$$

where $W^c=N\setminus W$ and $i(Y)=$ the number of isolated nodes of $\Gamma:Y$ which lie outside U.

COROLLARY 10. *Let $\alpha_k(\Gamma)=$ the number of partitions of the nodes of Γ into k connected blocks. Then*

$$a_k(H[\Gamma])=\alpha_k(\Gamma),$$

$$a_k(H[\pm\Gamma^\circ]) = \sum_{W\subseteq N} 2^{\#(W)-k}\alpha_k(\Gamma:W).$$

If $U\subseteq N$,

$$a_k(H[\pm\Gamma^U]) = \sum_W 2^{\#(W)-k}\alpha_k(\Gamma:W)$$

summed over those node sets W such that all the isolated nodes of $\Gamma:W^c$ *are in* U.

To prove the case $H=H[\pm\Gamma^U]$ we only have to look at the terms of highest degree in $\lambda^{i(N)}w_H(x,\lambda)$, for a_k is the coefficient of $x^{n-k}\lambda^k$. I should point out that $d(H)=i(N)$. You can satisfy yourself that there is a one-to-one correspondence between closed arc sets and the partitions of Γ enumerated by α, so the terms of highest degree in $w_{\Gamma:W}(x,\frac{1}{2}(\lambda-1))$ are $\alpha_k(\Gamma:W)x^{\#(W)-k}\lambda^k 2^{-k}$. The rest is easy.

The analog for f_k requires the numbers $\phi_l(\Gamma)$, defined by

$$\phi_l(\Gamma) = \sum_{T_l}(-1)^l\chi_{\Gamma/T_l}(-1)$$

summed over all the sets of arcs T_l which connect the nodes of Γ into l blocks.[10] Because a summand is 0 if T_l is not closed and is the number of regions in the arrangement $H[\Gamma/T_l]$ if T_l is closed, ϕ_l is positive. Given the definitions of ϕ_l and w_H and the fact that $f_k(H)$ is the coefficient of x^{n-k} in $(-1)^{n-i(N)}w_H(-x,-1)$, the deduction of Corollary 11 from Theorem 9 is merely a series of formal manipulations.

TABLE 1.

The numbers of flats and faces of some signed-graphic arrangements of planes, calculated by Corollaries 10 and 11 and the data in Fig. 9. The arrangements are $H[\pm\Gamma^U]$, where Γ is the graph in Fig. 9 and U is various.

(a) The arrangement $H[\pm\Gamma^\circ]$ (i.e., $U=N=\{1,2,3\}$). This arrangement is depicted in Fig. 6 (a).

k	0	1	2	3
Number of k-flats, a_k	1	5	5	1
Number of k-faces, f_k	1	10	24	16

(b) The arrangement $H[\pm\Gamma^{\{1,3\}}]$, depicted in Fig. 7 (a).

k	0	1	2	3
a_k	1	4	4	1
f_k	1	8	18	12

(c) The arrangement $H[\pm\Gamma^{\{1\}}]$, depicted in Fig. 8 (a).

k	0	1	2	3
a_k	0	1	3	1
f_k	0	1	6	6

(d) The arrangement $H[\pm\Gamma]$ (i.e., $U=\varnothing$), which consists of the two planes $x_1=\pm x_2$, meeting in the line $x_1=x_2=0$.

k	0	1	2	3
a_k	0	1	2	1
f_k	0	1	4	4

COROLLARY 11. *The face numbers of graphic and sign-symmetric arrangements are given by*

$$f_k(H[\Gamma]) = \phi_k(\Gamma),$$

$$f_k(H[\pm\Gamma^\circ]) = \sum_{W\subseteq N} 2^{\#(W)}\phi_k(\Gamma:W),$$

$$f_k(H[\pm\Gamma^U]) = \sum_{i=0}^{k}(-1)^i \sum_{W:\,i(W^c)=i} 2^{\#(W)}\phi_{k-i}(\Gamma:W).$$

When we are dealing with the root system arrangements A_{n-1}^*, B_n^*, and D_n^*, or with $D_n^{*(p)} = D_n^*$ plus p coordinate hyperplanes, Γ is the complete graph K_n. Every partition of the nodes is into connected blocks, so $\alpha_j(K_n) = S(n,j)$, the number of partitions of n objects into j blocks: the well-known Stirling number of the second kind. The contraction K_n/T_l, where T_l is closed, is a loop-free complete graph; its chromatic polynomial is therefore $(\lambda)_l$. Since there are $S(n,l)$ such sets T_l, $\phi_l(K_n) = S(n,l)l!$. Corollary 12 collects all the formulas we can now derive from Corollaries 10 and 11. The calculations are straightforward enough except that one needs the identity $S(n,k)/k = S(n-1,k) + S(n-1,k-1)/k$ in the evaluation of $f_k(D_n^{*(p)})$.

COROLLARY 12. *The flat and face numbers of A_{n-1}^*, B_n^*, and $D_n^{*(p)}$ are:*

$$a_k(A_{n-1}^*) = S(n,k),$$

$$f_k(A_{n-1}^*) = k!\,S(n,k),$$

$$a_k(B_n^*) = 2^{-k}\left[\sum_{j=k}^{n} 2^j\binom{n}{j}S(j,k)\right],$$

$$f_k(B_n^*) = k!\left[\sum_{j=k}^{n} 2^j\binom{n}{j}S(j,k)\right],$$

$$a_k(D_n^{*(p)}) = 2^{-k}\left[\sum_{j=k}^{n} 2^j\binom{n}{j}S(j,k) - (n-p)2^{n-1}S(n-1,k)\right],$$

$$f_k(D_n^{*(p)}) = k!\left[\sum_{j=k}^{n} 2^j\binom{n}{j}S(j,k) - (n-p)2^n\frac{1}{k}S(n,k)\right].$$

10. The End . . . That wraps up my discussion of subarrangements of the classical root system arrangements of hyperplanes. I believe I have made a case for the claim that just about anything about graphic and sign-symmetric arrangements can be reduced to ordinary graph theory. Arrangements which are neither graphic nor sign-symmetric can also be handled, but it takes a theory of signed graphs.[11]

A kind of question I have only touched on concerns the connection between the geometry of the arrangements based on an ordinary graph Γ and the combinatorics of Γ. Curtis Greene's discovery that regions of $H[\Gamma]$ correspond to acyclic orientations of Γ (see Section 7) suggests what one might find. As a matter of fact, regions of $H[\Sigma]$ correspond to acyclic orientations of the signed graph Σ. From that one can derive a combinatorial description of the way a region of $H[\Gamma]$ is subdivided when one passes to $H[\pm\Gamma^\circ]$. I believe there are more good problems along this line; certainly looking at geometry will lead to new ideas about graphs, and vice versa, as the connection between them is made increasingly strong.

Acknowledged (with pleasure). This work is a sea child, born on the warm, slow cruise of the *Rachael and Ebenezer* out of Rockland, Maine. To her and her people, many thanks. Thanks also to my colleagues at MIT, Joe Kung, Richard Stanley, and Jay Sulzberger, for their interest in and encouragement of signed graphs. To Jeff Lagarias and Bob Proctor for reading and editing the manuscript. To the NSF and SGPNR for their financial support. And foremost to Fred Supnick, who initiated me into combinatorial geometry, whose enthusiastic support sustained me mathematically through several hard, lean years, and who deserves many dedications.

Notes

1. For root systems and their connection to Lie algebras, see any book on Lie algebras, such as those of Wan, Adams, and Serre, or the excellent brief account of Veldkamp in [6, Section 3]. There are several definitions, all equivalent either to ours or to the slightly broader one adopted by Serre and Veldkamp (who call our root systems "reduced").

2. The characteristic polynomial is usually defined in terms of the lattice of flats; for this see [14]. The proof of Theorem 2 was found first by Winder, later (independently and in more generality) by me.

3. The first analysis of deletion and contraction invariants (called by Brylawski "Tutte-Grothendieck invariants") was Tutte's study of graphs. Later Brylawski extended the analysis to matroids (of which arrangements of hyperplanes are typical examples).

4. Signed graphs and some basic notions were invented by Harary. The connection between signed graphs and characteristic polynomials is implicit in Dowling's article, although only the case corresponding to B_n^* was discussed there.

5. For the characteristic (or rather, chromatic) polynomial of an arbitrary signed graph, see [16].

6. This is a classical proof. For various properties of chromatic polynomials, including a proof that they are polynomials, see Read's survey or a book on graph theory.

7. For a proof of this vector representation of a graph, see Theorem 2 in Section 9.5 of Welsh's book.

8. The Whitney polynomial was introduced in [14], there called the "Möbius" polynomial. Specialists will realize that "Whitney" is a better name, because the coefficients of $x^k\lambda^{n-k-d(H)}$ and $x^0\lambda^{n-k-d(H)}$ are the Whitney numbers W_k and w_k of the lattice of flats.

9. A proof from a different viewpoint appears in [16].

10. The ϕ_l have a combinatorial meaning for Γ. A theorem of Stanley's implies that the number of acyclic orientations of all contraction graphs Γ/T_l is $\phi_l(\Gamma)$.

11. See [15] and [16].

References

1. J. Frank Adams, Lectures on Lie Groups, Benjamin, New York, 1969.

2. Thomas H. Brylawski, A decomposition for combinatorial geometries, Trans. Amer. Math. Soc., 171 (1972) 235–282. [MR 46 #8869.]

3. T. A. Dowling, A class of geometric lattices based on finite groups, J. Combin. Theory Ser. B, 14 (1973) 61–86; Erratum, ibid., 15 (1973) 211. [MR 46 #7066.]

4. C. Greene, Acyclic orientations (Notes), in Higher Combinatorics, Martin Aigner, ed., D. Reidel, Dordrecht, 1977, pp. 65–68.

5. Frank Harary, On the notion of balance of a signed graph, Michigan Math. J., 2 (1953–1954) 143–146; Addendum, ibid., preceding p. 1. [MR 16, 733.]

6. M. Hazewinkel, W. Hesselink, D. Siersma, F. D. Veldkamp, The ubiquity of Coxeter-Dynkin diagrams (an introduction to the *A-D-E* problem), Nieuw Arch. Wisk. (3), 25 (1977) 257–307.

7. Ronald C. Read, An introduction to chromatic polynomials, J. Combin. Theory, 4 (1968) 52–71. [MR 37 # 104.]

8. Jean-Pierre Serre, Algèbres de Lie semi-simples complexes, Benjamin, New York, 1966.

9. Richard P. Stanley, Acyclic orientations of graphs, Discrete Math., 5 (1973) 171–178. [MR 47 #6537.]

10. W. T. Tutte, A ring in graph theory, Proc. Cambridge Philos. Soc., 43 (1947) 26–40. [MR 8, 284.]

11. Zhe-Xian Wan, Lie Algebras, Pergamon, Oxford, 1975.

12. D. J. A. Welsh, Matroid Theory, Academic Press, London, 1976.

13. R. O. Winder, Partitions of *N*-space by hyperplanes, SIAM J. Appl. Math., 14 (1966) 811–818. [MR 34 #8281.]

14. Thomas Zaslavsky, Facing Up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes, Mem. Amer. Math. Soc. No. 154, Amer. Math. Soc., Providence, R.I., 1975. [MR 50 #9603.]

15. ______, Signed graphs (submitted).

16. ______, Signed graph coloring (submitted).

PLANES, BIPLANES, AND THEIR CODES

CHESTER J. SALWACH

Department of Mathematics, Lafayette College, Easton, PA 18042

Design theory and algebraic coding theory have their origins in disparate fields of study: the statistical theory of the design of experiments and the theory of information transmission in electrical engineering. Yet each has enriched the other by providing tools capable of answering interesting and fundamental questions. A primary topic of this exposition will be the study of biplanes and projective planes, focusing on certain interesting relationships between these two types of combinatorial designs, relationships which are uncovered by exploiting the techniques of coding theory.

We shall set the stage in Section 1 by presenting a fairly complete introduction to the theory of designs, thereby putting both planes and biplanes in proper perspective. Section 2 presents the terminology of coding theory and a few of its tools. Then we examine both biplanes and planes from the viewpoint of coding theory. Additional design-theoretic tools are developed in Section 3, and in Section 4 much of the preceding is brought to bear in elucidating an elegant coding-theoretic link between planes and biplanes.

1. Introduction to Design Theory. A *t-design* (or *t-(v,k,λ) design*) on a v-set S (a finite set of size v whose elements are called *points*) is a collection, $\mathfrak{D}$, of k-subsets of S (called *blocks*) such that every t-subset of S is contained in precisely λ elements of $\mathfrak{D}$. We note that, although there are infinitely many nontrivial t-designs for each $t \leqslant 5$, no nontrivial designs are known for $t > 5$.

The collection consisting of all k-subsets of a v-set S trivially forms a *complete* k-$(v,k,1)$ design. The interesting case is when the collection is *incomplete* in the sense that not all k-subsets are present. In the statistical theory of the design of experiments, a primary root of design theory **[12]**, **[57]**, an incomplete 2-design is called a *balanced incomplete block design* (BIBD). The term "balanced" refers to the property that each pair of points, or treatments, as they are called, is contained in exactly λ blocks.

Below are three examples of (incomplete) t-designs. We shall often have occasion to refer to them. Each is related to the triangular figure presented in Fig. 1.

EXAMPLE 1.1. In this figure, we have seven points and seven geometrical objects (six line segments and one circle), each of which is incident with three points. Let us call these seven objects "lines." A 2-$(7,3,1)$ design is obtained by letting S consist of these seven points and letting a block consist of the three points incident with one of the seven lines. So, $\mathfrak{D} = \{B_i \mid 1 \leqslant i \leqslant 7\}$, where

$$B_1 = \{1,2,4\} \qquad B_2 = \{2,3,5\} \qquad B_3 = \{3,4,6\}$$
$$B_4 = \{4,5,7\} \qquad B_5 = \{5,6,1\} \qquad B_6 = \{6,7,2\}$$
$$B_7 = \{7,1,3\},$$

This paper is based on a series of invited lectures presented by the author at the Summer Conference on Combinatorics, co-sponsored by the Mathematics Department of Northern Michigan University and the Michigan Section of the MAA, and held at Northern Michigan University from July 17–21, 1978. The author is grateful to E. F. Assmus, Jr., and R. H. F. Denniston for reading an earlier version of this paper and making several useful suggestions.

Chester J. Salwach received his Ph.D. in Mathematics from Lehigh University, under E. F. Assmus, Jr., in 1976. Since that time he has been a member of the Mathematics Department of Lafayette College. His main areas of interest are block designs, coding theory, and combinatorial algorithms.—*Editors*

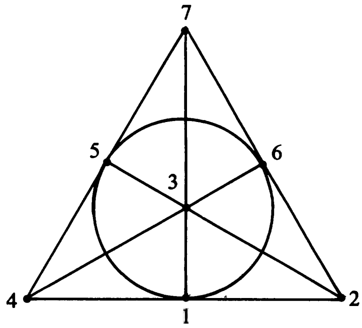


FIG. 1

and one can quickly check that every pair of points is contained in precisely one block.

EXAMPLE 1.2. We can also obtain a 2-(7, 4, 2) design from this figure. Again, let S be the set of seven points. Let a block consist of the four points not incident with one of the seven lines. Thus

$$B_1 = \{3,5,6,7\} \qquad B_2 = \{1,4,6,7\} \qquad B_3 = \{1,2,5,7\}$$
$$B_4 = \{1,2,3,6\} \qquad B_5 = \{2,3,4,7\} \qquad B_6 = \{1,3,4,5\}$$
$$B_7 = \{2,4,5,6\}.$$

Now, each pair of points is contained in precisely two blocks.

In each of the previous two examples, the number of blocks is the same as the number of points, each point is incident with the same number of blocks, this number being k, the number of points incident with a given block, and any two blocks meet in precisely λ points. Are these properties true in general? Let us examine the next example.

EXAMPLE 1.3. Let $S = \{1,2,3,4,5,6,7,8\}$ and let $\mathfrak{D}$ consist of the seven blocks of Example 1.2 and the seven blocks obtained by adjoining point "8" to each of the seven blocks of Example 1.1. One can verify that each 3-subset is contained in precisely one block, and so this collection of fourteen blocks forms a 3-(8, 4, 1) design on the set S.

In this last example, each point is incident with the same number of blocks, namely seven, but the other "properties" fail. Also, each pair of points is incident with precisely three blocks. Hence this 3-design is a 1-(8,4,7) and a 2-(8,4,3) design as well.

PROPOSITION 1.4. *If $(S, \mathfrak{D})$ is a t-(v,k,λ) design, then it is also a τ-(v,k,λ_τ) design for $0 \leqslant \tau \leqslant t$, with*

$$\lambda_\tau \binom{k-\tau}{t-\tau} = \lambda \binom{v-\tau}{t-\tau}.$$

Proof. Let X_l denote a set of l points. Let $\overline{X}_\tau$ be a given set of τ points and let $\bar{\lambda}_\tau$ be the number of blocks containing $\overline{X}_\tau$. Now, $\overline{X}_\tau$ is contained in $\bar{\lambda}_\tau$ blocks, each containing $k-\tau$ points other than those of $\overline{X}_\tau$, and so each of these $\bar{\lambda}_\tau$ blocks yields $\binom{k-\tau}{t-\tau}$ ways of extending $\overline{X}_\tau$ to a t-subset. Similarly, $\overline{X}_\tau$ can be extended to a t-subset in $\binom{v-\tau}{t-\tau}$ ways, and each resulting t-subset is contained in λ blocks. Thus, by counting the number of ordered pairs in the set

$$\left\{ (X_{t-\tau}, B) \mid B \in \mathfrak{D}, \overline{X}_\tau \dot{\cup} X_{t-\tau} \subset B \right\}$$

in two ways, we obtain

$$\binom{k-\tau}{t-\tau}\bar{\lambda}_\tau = \binom{v-\tau}{t-\tau}\lambda.$$

Solving for $\bar{\lambda}_\tau$, we see it depends only on the parameters of the design, and so $\bar{\lambda}_\tau$ is independent of the particular $\bar{X}_\tau$ chosen. Therefore $(S, \mathfrak{D})$ is a τ-design.

Thus for a t-(v, k, λ) design to exist, it is necessary that $\lambda\binom{v-\tau}{t-\tau}\Big/\binom{k-\tau}{t-\tau}$ be an integer for $0 \leqslant \tau \leqslant t$. This condition is *not* sufficient, however, since, for example, there does not exist a 2-(43, 7, 1) design, despite the fact that the numerical condition is satisfied for $\tau = 0$, 1, and 2.

Let us denote λ_0, the number of blocks, by b, and λ_1, the number of blocks containing any given point, by r. Setting $\tau = 0$ in Proposition 1.4 we obtain

$$b\binom{k}{t} = \lambda\binom{v}{t}.$$

So, in any 1-design we have

$$bk = rv,$$

since $\lambda = \lambda_1 = r$. We can obtain this last identity directly by counting the number of ordered pairs in the set

$$\{(P, B) \mid P \in S, B \in \mathfrak{D}, P \in B\}$$

in two ways. Also, in any 2-design we have, by taking $\tau = 1$ in Proposition 1.4,

$$r(k-1) = \lambda(v-1).$$

The *order* of a 2-design is $n = r - \lambda$. As we shall see, this parameter plays an important role in design theory.

The 3-(8, 4, 1) design of Example 1.3 is also a 2-(8, 4, 3) design, a 1-(8, 4, 7) design ($r = 7$), and a 0-(8, 4, 14) design ($b = 14$). Its order (as a 2-design) is $n = 7 - 3 = 4$.

In addition to trying to determine, for a given set of parameters, whether a design exists or not; that is, whether a t-(v, k, λ) design exists for a particular choice of t, v, k, and λ, it is also of importance to enumerate, if possible, all designs with a given set of parameters, once one has determined that at least one such design exists. We say two designs are *equivalent* or *isomorphic* if there exists a bijection of their point sets which transforms the blocks of one into those of the other. That is, two designs are equivalent if one can be obtained from the other simply by renaming the points and blocks, and one wishes, of course, to enumerate the isomorphism classes. The *automorphism group* of the design is

$$\{\sigma \in \mathrm{Sym}(S) \mid \sigma \cdot B \in \mathfrak{D}, \text{ for all } B \in \mathfrak{D}\},$$

where $\sigma \cdot B = \{\sigma(P) \mid P \in B\}$; that is, all permutations which leave the block structure invariant.

We next develop some tools and techniques which will help us to deal with the questions of existence and equivalence of designs.

The *incidence matrix* M of a design is a $b \times v$ matrix $(m_{B,P})$ where $m_{B,P} = 1$ if $P \in B$ and 0 otherwise. In the following, let I denote the identity matrix and J the matrix (of the appropriate size) whose every entry is 1.

LEMMA 1.5. *In any 2-design*:
(1) $MJ = kJ$.
(2) $JM = rJ$.
(3) $M^tM = (r-\lambda)I + \lambda J = nI + \lambda J$.
(4) $\mathrm{Det}(M^tM) = rk(r-\lambda)^{v-1} = rkn^{v-1}$.

Proof. (1) and (2) are obtained directly from the definitions of M and J.

(3) A column of M contains r 1's and two different columns have λ 1's in common.

(4)

$$M^tM=\begin{bmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & \cdots & \lambda \\ \vdots & \vdots & \ddots & \vdots \\ \lambda & \lambda & \cdots & r \end{bmatrix}.$$

Subtract the first column from each other column and obtain:

$$\begin{bmatrix} r & -n & \cdots & -n \\ \lambda & n & & 0 \\ \vdots & & \ddots & \\ \lambda & 0 & & n \end{bmatrix}.$$

Add each row to the first row and obtain:

$$\begin{bmatrix} r+\lambda(v-1) & & 0 \\ \lambda & n & \\ \vdots & & \ddots \\ \lambda & 0 & n \end{bmatrix}.$$

Therefore $\det(M^tM)=(r+\lambda(v-1))n^{v-1}=rkn^{v-1}$, since $r(k-1)=\lambda(v-1)$.

In each of our three examples there were at least as many blocks as there were points. Our next proposition, known as Fisher's Inequality, establishes this property for all interesting designs.

PROPOSITION 1.6. *If $(S,\mathfrak{D})$ is a t-design with $t\geqslant 2$ and $k\leqslant v-1$, then $b\geqslant v$.*

Proof. $(S,\mathfrak{D})$ is a 2-design. Recall $r(k-1)=\lambda(v-1)$. So, if $r=\lambda$, then $v=k$. Therefore $\det(M^tM)\neq 0$, and so M^tM is nonsingular, and hence of rank v. Thus $v=\mathrm{Rk}(M^tM)\leqslant \mathrm{Rk}(M)\leqslant b$, since M is a $b\times v$ matrix.

The hypothesis of a 2-design in Proposition 1.6 cannot be weakened, since if we let $S=\{1,2,3,4\}$ and $\mathfrak{D}=\{\{1,2\},\{3,4\}\}$, then $(S,\mathfrak{D})$ is a 1-(4,2,1) design. But $b=2<4=v$.

Let $(S,\mathfrak{D})$ be a t-(v,k,λ) design. Pick $P\in S$. Let $S_P=S-\{P\}$ and $\mathfrak{D}_P=\{B-\{P\}\mid B\in\mathfrak{D}, P\in B\}$. $(S_P,\mathfrak{D}_P)$ is called the *contraction* of $(S,\mathfrak{D})$ at P and it is a $(t-1)$-$(v-1,k-1,\lambda)$ design. Note that the "r" for $(S,\mathfrak{D})$ is the "b" for $(S_P,\mathfrak{D}_P)$.

Thus the existence of a single t-design establishes the existence of several possibly nonisomorphic $(t-1)$-designs. In Example 1.3, by contracting on point "8", we obtain the design of Example 1.1. By contracting on any other point, we obtain a design isomorphic to that of Example 1.1.

We immediately exploit this notion of contraction in proving the next proposition.

PROPOSITION 1.7. *If $(S,\mathfrak{D})$ is an incomplete t-(v,k,λ) design with $t\geqslant 3$, then $b>v$.*

Proof. We know $b\geqslant v$ by Proposition 1.6. By contracting on a point P of S, we obtain a $(t-1)$-$(v-1,k-1,\lambda)$ design. Since $t-1\geqslant 2$, we have, by Proposition 1.6, $r=b_{S_P}\geqslant v-1$, where b_{S_P} denotes the number of blocks in the $(t-1)$-design. Say $b=v$. Then, since $vr=bk$, we get $k=r\geqslant v-1$. If $k=v$, we have the complete v-$(v,v,1)$ design, and if $k=v-1$, we have the complete $(v-1)$-$(v,v-1,1)$ design. Thus $b>v$.

There is a special name for the interesting and important class of 2-designs for which the number of blocks and points is the same. A *projective* (or *symmetric*) design is one for which $t>1$ and $b=v$ (and so $r=k$ as well). Thus, except for some complete designs, a projective design is a 2-design, and no more; that is, it cannot be a t-design with $t\geqslant 3$. We shall therefore suppress the "2" and speak of projective (v,k,λ) designs.

The *dual* of a design $(S,\mathfrak{D})$, denoted by $(S^d,\mathfrak{D}^d)$, is obtained by interchanging the notions of point and block. If $(S,\mathfrak{D})$ is a 1-(v,k,r) design, then $(S^d,\mathfrak{D}^d)$ is a 1-(b,r,k) design. Can both the design and its dual be better than 1-designs? If the design is no better than a 1-design, then clearly no. If it is an incomplete 3-design, then Proposition 1.7 implies that the dual design will have fewer blocks than points, and so Fisher's Inequality (Proposition 1.6) provides us with the negative answer. The next proposition provides us with an affirmative answer in the case of the projective design.

PROPOSITION 1.8. *If* $(S,\mathfrak{D})$ *is a projective* (v,k,λ) *design, then* $(S^d,\mathfrak{D}^d)$ *is as well; that is, any two distinct blocks of* $\mathfrak{D}$ *have precisely* λ *points in common.*

Proof. As we saw in the proof of Fisher's Inequality, $v\leqslant \mathrm{Rk}(M)\leqslant b$. Since $b=v$ for a projective design, $\mathrm{Rk}(M)=v$. So M^{-1} exists and by (3) of Lemma 1.5 we obtain $M^t=[(r-\lambda)I+\lambda J]M^{-1}$. Since $r=k$, parts (1) and (2) of Lemma 1.5 imply $MJ=JM$. Therefore $MM^t=M^tM=(r-\lambda)I+\lambda J$. Hence any two distinct blocks of $\mathfrak{D}$ have precisely λ points in common.

In the remainder of this exposition we shall be concerned primarily with projective designs for which $\lambda=1$ or 2. A (projective) *plane* is a projective design with $\lambda=1$; that is, an $(n^2+n+1,n+1,1)$ design, where $n=k-1$ is the order. The blocks of a plane are commonly called *lines*. A *biplane* is a projective design with $\lambda=2$; that is, a $(\frac{1}{2}(n^2+3n+4),n+2,2)$ design, where $n=k-2$ is the order. The parameters of these projective designs are obtained by using the identity $k(k-1)=\lambda(v-1)$, solving for v, and then expressing each parameter in terms of the order.

The design of Example 1.1 is a plane of order 2 and the design of Example 1.2 is a biplane of order 2.

Only finitely many projective designs are known to exist for any fixed $\lambda>1$, and it is conjectured that indeed there exist but finitely many projective designs for any fixed $\lambda>1$. However, an infinite class of planes, called Desarguesian planes of order n, are constructed from finite fields with $n=p^a$ elements where p is a prime and a is a positive integer [**23**], [**29**], [**53**]. We shall indicate a construction in Section 2.

We shall not call the complete 2-(3,2,1) design a plane of order 1. Normally, one of the defining properties of a projective plane is that it contain a set of four triply noncollinear points. Each of our planes will satisfy this property.

The following chart exhibits the current state of affairs for biplanes.

n	1	2	3	4	5	6	7	8	9	10	11
number	1	1	1	3	0	0	4	0	$\geqslant 4$	0	$\geqslant 2$

It indicates that the biplanes of orders 1 through 3 are unique. Also, there exist precisely three biplanes of order 4 [**33**], four of order 7 [**49**], at least four of order 9 [**50**], and at least two of order 11 [**2**]. The two of order 11 and two of the biplanes of order 7 are duals of each other and the rest are self-dual (they are isomorphic to their duals). We shall discuss many of these biplanes in more detail in the succeeding sections.

The following theorem presents the strongest known necessary condition for the existence of a projective design. It is the fundamental theorem of the subject. It was first proved by Bruck and Ryser [**14**] for the case of planes (see Corollary), and then extended by Chowla and Ryser [**17**] to cover an arbitrary projective design.

THEOREM 1.9. *If* $(S, \mathfrak{D})$ *is a projective* (v, k, λ) *design of order* $n = k - \lambda$, *then*

(1) v *even implies* n *is a square*,

(2) v *odd implies* $x^2 = ny^2 + (-1)^{(1/2)(v-1)}\lambda z^2$ *has a solution in integers* x, y, *and* z, *not all* 0.

Proof. By Lemma 1.5, $\det(M) = \det(M^t) = (\det(M^t M))^{1/2} = (rk(r-\lambda)^{v-1})^{1/2} = kn^{(1/2)(v-1)}$. Hence, if v is even, then $n^{1/2}$ must be an integer.

The proof of (2) involves sophisticated number theoretic techniques, and we shall not present it here.

It is conceivable that these necessary conditions are even sufficient, since the nonexistence of a design with parameters satisfying these conditions has never been proved.

Since a biplane of order 5 would be a (22, 7, 2) projective design, part (1) of the theorem guarantees its nonexistence.

COROLLARY. *If* $n \equiv 1$ *or* 2 (mod 4) *and* n *cannot be expressed as the sum of two integral squares, then there does not exist a plane of order* n.

Since $6 \equiv 2 \pmod 4$ and 6 cannot be expressed as the sum of two integral squares, there does not exist a plane of order 6.

It is possible to prove by means of algebraic coding theoretic techniques that there exists no plane of order $n \equiv 6 \pmod 8$ and no biplane of order $n \equiv 5 \pmod 8$ [**3**].

There have been many attempts to construct the putative plane of order 10 (see [**40**]), which is the smallest unsettled case, and we shall present one such attempt in Section 4. Such a plane would, of course, be a projective (111, 11, 1) design, and the search space is simply much too large to exhaust all possibilities by electronic computation on a computer. It is conjectured by some, however, that every plane is of prime power order. Certainly every known plane is of prime power order.

In addition to the existence question, the question of enumeration is also of interest; that is: How many nonisomorphic planes exist for a given order? It is known that there exist planes other than those constructed from the finite fields. There exist unique planes of orders 2, 3, 4, 5, 7, and 8 (see [**28**]), but there exist at least four nonisomorphic planes of order 9 (see [**27**] and [**53**, p. 72]), and at least five nonisomorphic planes of order 16 [**36**], [**37**]. Of these, the non-Desarguesian planes are constructed from Veblen-Webberburn systems (quasifields) and semifields (nonassociative division rings).

Inverse to the operation of contracting a design, which we dealt with previously, is the notion of possibly *extending* a design. A $(t+1)$-design $(S, \mathfrak{D})$ is said to be an *extension* of a t-design $(S', \mathfrak{D}')$ if $(S', \mathfrak{D}')$ is isomorphic to $(S_P, \mathfrak{D}_P)$, the contraction of $(S, \mathfrak{D})$ at P, for some point P in S.

The 3-(8,4,1) design of Example 1.3 is the unique extension of the plane of order 2, the 2-(7,3,1) design of Example 1.1. This 3-design cannot be extended further (see Lemma 1.10 below).

LEMMA 1.10. *If a* t-(v, k, λ) *design is extendible, then* $k+1$ *divides* $b(v+1)$.

Proof. Denote a parameter of the extension by placing a $^-$ above the letter. The extension, if it exists, is a $\bar{t}-(\bar{v}, \bar{k}, \bar{\lambda}) = (t+1)-(v+1, k+1, \lambda)$ design with $\bar{r} = b$. Thus $\bar{b}\bar{k} = \bar{v}\bar{r}$ implies $\bar{b}(k+1) = (v+1)b$.

THEOREM 1.11. *If a plane of order* n *is extendible, then* $n = 2, 4$, *or* 10 (*see* [**31**]).

Proof. A plane of order n is a 2-$(n^2+n+1, n+1, 1)$ design and its extension would be a 3-$(n^2+n+2, n+2, 1)$ design. By Lemma 1.10, $b(v+1) \equiv 0 \pmod{k+1}$. That is,

$$(n^2+n+1)(n^2+n+2) \equiv 0 \pmod{n+2}.$$

Since $n \equiv -2 \pmod{n+2}$, we obtain

$$(n^2+n+1)(n^2+n+2) \equiv (4-2+1)(4-2+2) = 12 \equiv 0 \pmod{n+2},$$

thereby implying that $n+2$ divide 12. Therefore

$$n+2 \in \{1,2,3,4,6,12\}$$

and so

$$n \in \{-1,0,1,2,4,10\}.$$

Thus $n = 2, 4,$ or 10.

Thus if the plane of order 10 does exist, it might have an extension. The plane of order 4, a 2-(21, 5, 1) design, can be uniquely extended three times to a 5-(24, 8, 1) design. By Lemma 1.10, however, it can be extended no further.

We conclude this section with a method of constructing infinitely many projective designs with $\lambda > 1$ via the planes (of course, but finitely many for each fixed λ). The next lemma (see **[1]** and **[41]**) prepares the way.

LEMMA 1.12. *For a t-(v,k,λ) design, the number of blocks meeting an i-subset of S in precisely j points, where $j \leqslant i \leqslant t$, depends only on i, j, and the parameters of the design.*

Proof. Let N_j equal the number of blocks meeting a given i-subset, say X_i, in exactly j points. Then we obtain the following system of $i+1$ equations in $i+1$ unknowns:

$$\begin{aligned}
N_0 + N_1 + N_2 + N_3 + \cdots + N_i &= \lambda_0 = b\\
N_1 + 2N_2 + 3N_3 + \cdots + iN_i &= i\lambda_1 = ir\\
N_2 + \binom{3}{2}N_3 + \cdots + \binom{i}{2}N_i &= \binom{i}{2}\lambda_2\\
&\vdots\\
N_i &= \binom{i}{i}\lambda_i.
\end{aligned}$$

Recall that, by virtue of Proposition 1.4, a t-design is also a τ-design for $0 \leqslant \tau \leqslant t$. The second equation is obtained by counting the number of ordered pairs in the set

$$\{(P,B) \mid P \in X_i, P \in B\}$$

in two ways. The third is obtained similarly by counting the elements of

$$\{(\{P_1,P_2\},B) \mid \{P_1,P_2\} \subset X_i, \{P_1,P_2\} \subset B\}$$

in two ways. And so forth. This system can be solved simultaneously by starting with the last equation and proceeding to the first, and the resulting N_j's are therefore seen to be independent of the particular X_i chosen.

PROPOSITION 1.13. *Let $(S,\mathfrak{D})$ be a projective (v,k,λ) design, and let $\mathfrak{D}^c = \{S-B \mid B \in \mathfrak{D}\}$; then $(S,\mathfrak{D}^c)$, called the complement of $(S,\mathfrak{D})$, is a projective*

$$(v, v-k, (v-k)(v-k-1)/(v-1)) \text{ design}.$$

Proof. If it is a 2-design, it will clearly be a projective design since the number of blocks and points is the same as for the original design. Now, it is a 2-design since the number of blocks of $(S,\mathfrak{D})$ not meeting a 2-subset of S is $\lambda_c = N_0 = v - N_1 - N_2$ (by Lemma 1.12 with $i = t = 2$), where λ_c is the "λ" for the complementary design. Using the equations derived in the proof of Lemma 1.12 and solving for N_2, N_1, and N_0 we obtain:

$$\begin{aligned}
N_2 &= \lambda,\\
N_1 &= 2k - 2N_2 = 2k - 2\lambda,
\end{aligned}$$

and so

$$\begin{aligned}\lambda_c = N_0 &= v + \lambda - 2k \\ &= v + \frac{k(k-1)}{v-1} - 2k, \quad \text{since } \lambda\binom{v}{2} = b\binom{k}{2}, \\ &= (v-k)(v-k-1)/(v-1).\end{aligned}$$

Thus the parameters of the design are correct.

The complementary design of the plane of Example 1.1 is the biplane of Example 1.2. In this manner, we can construct infinitely many projective designs with $\lambda > 1$ as the complements of projective planes.

The reader may wish to consult Biggs **[10]**, Cameron and van Lint **[16]**, Dembowski **[18]**, Hall **[24]**, and Ryser **[45]** for additional material on design theory.

2. Codes and Ranks of Incidence Matrices. Coding theory has its roots in the problem of transmitting and correctly recovering information symbols which are sent over a noisy channel (see **[9]**). Surprisingly, there is great interplay between coding theory and design theory. We shall exploit this interrelationship and use certain tools of coding theory to answer some questions in the theory of designs in this and later sections. We first present some basic notions of algebraic coding theory.

A *linear* (n, l) *code* A over the field F_q with q elements, q a prime power, is an l-dimensional subspace of F_q^n, the vector space of all n-tuples with entries from F_q. The elements of A are called *code words* or *vectors*. For $a \in A$, the *weight* of a is

$$\mathrm{wgt}(a) = |\{i \mid a_i \neq 0\}|,$$

where $a = (a_1, a_2, \ldots, a_n)$, and $|X|$ denotes the number of elements in the set X. The *minimum weight* of A is

$$d(A) = \mathrm{Min}\{\mathrm{wgt}(a) \mid a \in A, a \neq \mathbf{0}\}.$$

The *orthogonal* of A is

$$A^\perp = \{b \in F_q^n \mid a \cdot b = 0, \forall a \in A\},$$

where $a \cdot b$ is the usual dot product. If A is an (n, l) code over F_q, then $A^\perp$, its orthogonal, is an $(n, n-l)$ code over F_q (see **[39**, p. 26**]**). Note that since the arithmetic is performed in F_q, the dot product of a nonzero vector with itself might very well be zero. So we shall say A is *self-orthogonal* if $A \subseteq A^\perp$ and *self-dual* if $A = A^\perp$.

If one defines the *distance* between two code vectors a and b of A by $\mathrm{dist}(a, b) = \mathrm{wgt}(a - b)$, then this distance function is a metric for the code A.

For additional information on the subject of codes, see Assmus and Mattson **[4]**, Berlekamp **[9]**, Cameron and van Lint **[16]**, MacWilliams and Sloane **[39]**, Peterson and Weldon **[43]**, and van Lint **[55]**.

Let M be the incidence matrix of a 2-design $(S, \mathfrak{D})$ and let p be a prime. The *p-rank of M*, denoted by $\mathrm{Rk}_p(M)$, is the dimension over F_p of the rowspace of M; that is, the rows of M generate a $(v, \mathrm{Rk}_p(M))$ code over F_p, which we shall denote by A_p or $SP_p(M)$. For $a \in A_p$, the *support of a* is defined to be

$$\mathrm{supp}(a) = \{P_i \mid P_i \in S, a_i \neq 0\}.$$

The next theorem, primarily due to Hamada **[29]**, is one illustration of the importance of the parameter n, the order of a design.

Let us denote $(1, 1, \ldots, 1)$, the all-one vector, by $\mathbf{1}$.

THEOREM 2.1. *Let M be the incidence matrix of a* 2-(v, k, λ) *design, then* $\mathrm{Rk}_p(M)$ *depends only on the parameters of the design unless p divides n, the order (in which case* $\mathrm{Rk}_p(M)$ *may depend*

upon the block structure of the design). In particular:

(1) *If* $p \nmid rn$, *then* $\mathrm{Rk}_p(M)=v$.

(2) *If* $p \mid r$ *and* $p \nmid kn$, *then* $\mathrm{Rk}_p(M)=v$.

(3) *If* $p \mid r$, $p \mid k$, *and* $p \nmid n$, *then* $\mathrm{Rk}_p(M)=v-1$ *and* $A_p=\langle \mathbf{1}\rangle^{\perp}$, *the orthogonal of the code generated by the all-one vector.*

Proof. Recall $n=r-\lambda$.

(1) Let w be the vector sum of all the rows of M; then $w=(r,r,\dots,r)$. Denote by s_i the vector sum of those rows of M which have a 1 in column i; that is, s_i is constructed from the blocks containing point P_i. Then $s_i=(\lambda,\dots,\lambda,r,\lambda,\dots,\lambda)$, where r inhabits the ith position. Thus

$$rs_i-\lambda w=(0,\dots,0,rn,0,\dots,0)\not\equiv\mathbf{0}(\mathrm{mod}\,p),$$

thereby yielding a collection of v linearly independent vectors. Hence $\mathrm{Rk}_p(M)=v$.

(2) With s_i as before, consider

$$s_i-s_v=(0,\dots,0,n,0,\dots,0,-n)\not\equiv\mathbf{0}(\mathrm{mod}\,p),$$

for $1\leqslant i\leqslant v-1$. Each s_i-s_v is in $\langle\mathbf{1}\rangle^{\perp}$, which is $(v-1)$-dimensional, and $\{s_i-s_v \mid 1\leqslant i\leqslant v-1\}$ is linearly independent. Hence the code spanned by $\{s_i-s_v \mid 1\leqslant i\leqslant v-1\}$ equals $\langle\mathbf{1}\rangle^{\perp}$. Thus $\langle\mathbf{1}\rangle^{\perp}$ is a subspace of A_p and $\mathrm{Rk}_p(M)\geqslant v-1$. Now each row of M is necessarily in A_p, but cannot be in $\langle\mathbf{1}\rangle^{\perp}$ since $p\nmid k$. Therefore $\mathrm{Rk}_p(M)=v$.

(3) From part (2), we saw that $\mathrm{Rk}_p(M)\geqslant v-1$. If $p\mid k$, then $A_p\subseteq\langle\mathbf{1}\rangle^{\perp}$, since each row of M has k 1's, and so $\mathrm{Rk}_p(M)\leqslant v-1$. Hence $\mathrm{Rk}_p(M)=v-1$ and $A_p=\langle\mathbf{1}\rangle^{\perp}$.

Thus only when p divides the order is the notion of p-rank possibly useful in differentiating among designs with the same parameters. (If the incidence matrices of two designs have different p-ranks, then they are necessarily nonisomorphic.) For the three biplanes of order 4, only the 2-rank is of interest in possibly distinguishing them. In fact, the 2-ranks of the three are 6, 7, and 8, respectively.

In the case of a projective design, we can say a little more about the situation in which the prime of interest divides the order of the design.

PROPOSITION 2.2. *Let M be the incidence matrix of a projective (v,k,λ) design. If $p\mid n$ but $p^2 \nmid nk$, then* $\mathrm{Rk}_p(M)\geqslant\frac{1}{2}(v+1)$.

Proof. By the theory of elementary divisors, there exist unimodular matrices (matrices with integral coefficients and with a determinant of 1) P and Q such that PMQ is the diagonal matrix diag $\{d_1,d_2,\dots,d_v\}$, where $d_1\mid d_2\mid\cdots\mid d_v$. The product of the elementary divisors is $d_1d_2\cdots d_v$ $=\det(PMQ)=\det(M)=kn^{(1/2)(v-1)}$ (by part (4) of Lemma 1.5 with $r=k$). Thus p divides $\det(M)$ a total of $\frac{1}{2}(v-1)$ times. Hence $\mathrm{Rk}_p(M)=\mathrm{Rk}_p(PMQ)\geqslant v-\frac{1}{2}(v-1)=\frac{1}{2}(v+1)$ since at most $\frac{1}{2}(v-1)$ of the latter d_i's have p as a factor.

In the case of planes or biplanes of prime order, our next two propositions provide us with an answer in the situation in which the prime of interest is the order of the design.

PROPOSITION 2.3. *If M is the incidence matrix of a plane, then $d_v=(n+1)n$. Moreover, if $n=p$, then* $\mathrm{Rk}_p(M)=\frac{1}{2}(v+1)$.

Proof. Now $d_v=\Delta_v/\Delta_{v-1}$, where

$$\Delta_i=\gcd\{\det(M_i) \mid M_i \text{ is an } i\times i \text{ submatrix of } M\}.$$

A standard calculation, such as the one involved in proving part (4) of Lemma 1.5 (but much longer), yields

$$\Delta_{v-1}=\gcd\{n^{(1/2)(v-1)},n^{(1/2)(v-3)}\}.$$

Since $\Delta_v = \det(M) = kn^{(1/2)(v-1)} = (n+1)n^{(1/2)(v-1)}$, we obtain $d_v = (n+1)n$. If $n=p$, then, by Proposition 2.2, $\mathrm{Rk}_p(M) \geqslant v - \frac{1}{2}(v-1) = \frac{1}{2}(v+1)$. Also, the last $\frac{1}{2}(v-1)$ d_i's have p as a factor. Hence $\mathrm{Rk}_p(M) = \frac{1}{2}(v+1)$.

PROPOSITION 2.4. *If M is the incidence matrix of a biplane, then $d_v = \frac{1}{2}(n+2)n$ if 2 divides n and $(n+2)n$ otherwise. Moreover, if $2 \neq n = p$, then* $\mathrm{Rk}_p(M) = \frac{1}{2}(v+1)$.

Proof. Another standard, but tedious, calculation yields

$$\Delta_{v-1} = \gcd\{n^{(1/2)(v-1)}, 2n^{(1/2)(v-3)}\},$$

from which the results follow as before.

Additional information on the preceding three propositions can be found in **[42]**, **[46]**, and **[48]**.

Thus each of the four biplanes of order 7 must have a 7-rank of 19, and so the concept of p-rank cannot aid us in distinguishing these four biplanes. (The 3-rank for each is 36 and the p-rank where p is other than 3 or 7 must be 37 by virtue of Theorem 2.1.) We shall see in Section 3 that the notion of λ-*chain* is useful in distinguishing them.

Let $q = p^a$ be a prime power. A $\mathrm{PG}(t,q):\mu$ 2-design (projective geometry design) is one in which the points are the 1-dimensional subspaces of F_q^{t+1} and the blocks are the $(\mu + 1)$-dimensional subspaces of F_q^{t+1}.

$\mathrm{PG}(2,q):1$ is the Desarguesian plane of order q, and in this manner we obtain infinitely many planes, one of each prime power order. The plane of Example 1.1 is $\mathrm{PG}(2,2):1$.

For these particular designs, the question of p-rank has been completely answered. If $p \nmid n$, then of course consult Theorem 2.1, whereas if $p \mid n$, see below.

MacWilliams and Mann **[38]** obtained

$$\mathrm{Rk}_p(M) = \binom{p+1}{2}^a + 1,$$

where M is the incidence matrix for $\mathrm{PG}(2,q):1$, the Desarguesian plane of order q. Smith **[52]**, and Goethals and Delsarte **[20]** obtained

$$\mathrm{Rk}_p(M) = \binom{p+t-1}{t}^a + 1,$$

where M is the incidence matrix for $\mathrm{PG}(t,q):t-1$, the projective design of points and hyperplanes. Hamada (see **[29]**) obtained the formula for $\mathrm{Rk}_p(M)$, where M is the incidence matrix for $\mathrm{PG}(t,q):\mu$, the general projective geometry design. However, it is too complicated to present here.

Hamada **[29]** has conjectured that $\mathrm{Rk}_p(\mathrm{PG}(t,q):\mu) \leqslant \mathrm{Rk}_p(M)$, where M is the incidence matrix for a design with the same parameters as those of $\mathrm{PG}(t,q):\mu$, with equality if and only if the design is $\mathrm{PG}(t,q):\mu$. Hamada and Ohmori **[30]** have proved the conjecture in the case of $q=2$ and $\mu=t-1$, but their proof does not appear to be generalizable.

Of course, all known examples bear out Hamada's conjecture. In the case of the four known planes of order 9, the 3-rank of the Desarguesian one is of course 37, and the other three known planes **[53]**, constructed from quasifields, each have a 3-rank of 41 **[46]**. The author has studied five planes of order 16. The Desarguesian plane has a 2-rank of 82, the three planes constructed from quasifields presented in **[36]** have a 2-rank of 98 **[48]**, and the fifth plane **[37]**, also constructible from a quasifield, has a 2-rank of 122 **[48]**. These interesting results for the non-Desarguesian planes of orders 9 and 16 certainly warrant further study.

Since every plane of order p has, by virtue of Proposition 2.3, a p-rank of $\frac{1}{2}(v+1) = \binom{p+1}{2} + 1$, Hamada's conjecture would imply the famous conjecture that the only plane of

order p is the Desarguesian one, PG$(2, p)$:1.

The following theorem examines, in more detail, the code A_p generated by the incidence matrix of a plane of order n, where we take p to be a divisor of n, the interesting situation discussed above. The proof of this theorem for Desarguesian planes can be found in **[21]**, and for an arbitrary plane see **[46]**, **[47]**. Additional information on this subject is also contained in chapter 11 of **[16]**.

THEOREM 2.5. *Let M be the incidence matrix for a plane of order n and let A_p denote, as before, the code generated by the rows of M over F_p, where p divides n. Then*:

(1) *The minimum weight of A_p is $n + 1$; that is, $d(A_p) = k$.*

(2) $\mathrm{Wgt}(a) = k$ *for $a \in A_p$ if and only if $a = \gamma l$ for some nonzero $\gamma \in F_p$ and l a row of M.*

The support of a row of M is, of course, a line of the plane. Amazingly, one thus recovers the plane from the $(\bmod p)$ span of its incidence matrix simply by taking the supports of the minimum weight vectors. The entire plane can therefore be reconstructed from a basis for the code. This theorem can also be used, of course, in testing whether a collection of k-subsets can be enlarged to form the collection of lines of a plane.

However, neither part of Theorem 2.5 holds true for biplanes in general, as is illustrated below in the case of the three biplanes of order 4, the projective $(16, 6, 2)$ designs. We first indicate coding theoretic constructions for these three biplanes.

Let $\mathrm{SP}(\mathfrak{B}_i)$ denote the $(16, i)$ code over F_2 generated by the rows of the incidence matrix M for the biplane $\mathfrak{B}_i$, where $i = 6$, 7, or 8 is the 2-rank of M.

Construct a 5×16 matrix N as follows. The first four rows are formed by taking the sixteen columns to be the sixteen vectors in F_2^4. The last row consists entirely of 1's. The (mod 2) span of N is the $(16, 5)$ first-order Reed-Muller code, and we denote it by $H^\perp$. The orthogonal of this code, which we denote by H, is the extended binary $(16, 11)$ Hamming code. Since each row of N contains an even number of 1's and the dot product of any two distinct rows is even, we see that $H^\perp \subset H$, and so $H^\perp$ is self-orthogonal.

The weight distribution for $H^\perp$ (where nx^i means n code vectors of weight i) is

$$x^0 + 30x^8 + x^{16}$$

and the weight distribution for H is

$$x^0 + 140x^4 + 448x^6 + 870x^8 + 448x^{10} + 140x^{12} + x^{16}.$$

PROPOSITION 2.6. *Let $F = F_2$.*

(1) *Let v be any weight-6 vector of H. Let $B_6 = H^\perp + Fv$. Then the weight distribution for B_6 is*

$$x^0 + 16x^6 + 30x^8 + 16x^{10} + x^{16}.$$

Moreover, $B_6 = \mathrm{SP}(\mathfrak{B}_6)$ and the 448 weight-6 vectors of H split naturally into 28 disjoint $\mathfrak{B}_6$'s.

(2) *Let v be a weight-6 vector in $(\mathrm{SP}(\mathfrak{B}_6))^\perp - \mathrm{SP}(\mathfrak{B}_6)$ and let $B_7 = \mathrm{SP}(\mathfrak{B}_6) + Fv$. Then the weight distribution for B_7 is*

$$x^0 + 4x^4 + 32x^6 + 54x^8 + 32x^{10} + 4x^{12} + x^{16}.$$

Moreover, $B_7 = \mathrm{SP}(\mathfrak{B}_7)$ and the 32 weight-6 vectors of B_7 split into two disjoint $\mathfrak{B}_6$'s in exactly four ways and two disjoint $\mathfrak{B}_7$'s in exactly four ways, thereby yielding the eight $\mathfrak{B}_6$'s and the eight $\mathfrak{B}_7$'s which can be formed from the 32 weight-6 vectors of B_7.

(3) *Let v be a weight-6 vector in $(\mathrm{SP}(\mathfrak{B}_7))^\perp - \mathrm{SP}(\mathfrak{B}_7)$ and let $B_8 = \mathrm{SP}(\mathfrak{B}_7) + Fv$. Then the weight distribution for B_8 is*

$$x^0 + 12x^4 + 64x^6 + 102x^8 + 64x^{10} + 12x^{12} + x^{16}.$$

Moreover, $B_8 = \mathrm{SP}(\mathfrak{B}_8)$ and the 64 weight-6 vectors of B_8 split into four disjoint $\mathfrak{B}_6$'s, $\mathfrak{B}_7$'s, and $\mathfrak{B}_8$'s. Also, there exist precisely 96 $\mathfrak{B}_6$'s, 288 $\mathfrak{B}_7$'s and 192 $\mathfrak{B}_8$'s in B_8.

We here present a proof of part (1) of this proposition. Proofs of the other parts, along with a

wealth of additional information on these three biplanes, can be found in [7] and [**48**].

Proof. Now $B_6 \subset B_6^{\perp}$, since v is in H and $H^{\perp} \subset H$. Let $a \in H^{\perp}$ be a vector of weight-8. Since $a \cdot v \in \{0,2,4,6\}$, we obtain $\mathrm{wgt}(a+v) \in \{14,10,6,2\}$. But possibilities 2 and 14 are impossible since $B_6 \subset H$ which has no vectors of weights 2 or 14. Since the all-one vector is in $H^{\perp}$, the number of vectors of weights 6 and 10 must be the same. Thus B_6 has the desired weight distribution.

Let b_i and b_j be any two distinct weight-6 vectors in B_6. Now $b_i \cdot b_j \in \{0,2,4\}$ thereby yielding $\mathrm{wgt}(b_i + b_j) \in \{12,8,4\}$, but B_6 has no vectors of weights 4 or 12. Hence

$$|\mathrm{supp}(b_i) \cap \mathrm{supp}(b_j)| = 2,$$

and so the supports of the 16 weight-6 vectors of B_6 form the blocks of a biplane (see Proposition 1.8) with a 2-rank of 6; that is, $B_6 = \mathrm{SP}(\mathfrak{B}_6)$. Since each weight-6 vector of H selects 15 others which together determine the 16 blocks of the biplane, the $488 = 16 \cdot 28$ weight-6 vectors of H split naturally into 28 disjoint $\mathfrak{B}_6$'s.

We know of no other such "nesting" of designs. By Theorem 2.5, such a nesting is clearly impossible for planes of the same order, since the supports of the minimum weight code vectors are precisely the lines of the plane. Of the three biplanes of order 4, only $\mathfrak{B}_6$ enjoys this property.

3. λ-chains, Difference Sets, and Ovals. In this section we present various notions which are useful in the characterization and construction of projective designs, particularly planes and biplanes. We shall make use of these ideas in Section 4.

We first present Hussain's manner of describing a biplane in terms of λ-chains.

1. Choose a block, call it the *indexing block*, and index its points from 1 to k.

2. Index each remaining block by the unique 2-subset in which it meets the indexing block.

3. Index each point P not incident with the indexing block by the collection of k 2-subsets which index the blocks that contain P. (Each point incident with the indexing block is contained in two of these 2-subsets.)

4. If the 2-subsets $\{a_1,a_2\},\{a_2,a_3\},\ldots,\{a_{n-1},a_n\},\{a_n,a_1\}$ index a collection of blocks incident with P, we express this by writing $(a_1,a_2,\ldots,a_{n-1},a_n)$. Each cycle must be at least of length 3. The collection of cycles for P is called the λ-*chain*, or simply *chain*, for P.

A λ-chain may also be viewed as a graph on the points of the indexing block where the points become the vertices and the 2-subsets become the edges. So each λ-chain becomes a graph of valency 2 which is a disjoint union of polygons.

We call a λ-chain consisting of cycles containing $c_1, c_2, \ldots, c_m$ letters, respectively, a type $(c_1 - c_2 - \cdots - c_m)$-chain. The number of distinct types of λ-chains which are possible for a biplane equals the number of partitions of k where each summand is at least 3. The *chain structure* of the biplane for a given indexing block is simply the number of chains of each type. Of course, $v-k$ chains comprise the chain structure for each choice of indexing block. These chains can be systematically examined by realizing that each pair of 2-subsets $\{1,i\}$ and $\{1,j\}$, where $2 \leqslant i < j \leqslant k$, must be associated with precisely one λ-chain (the blocks indexed by these 2-subsets meet in point "1" and in a point not incident with the indexing block), thereby yielding $v-k=\binom{k-1}{2}$.

Unless the automorphism group of the biplane is transitive on its blocks, the chain structure may depend on the choice of the indexing block. Two biplanes are necessarily nonisomorphic if the two collections of chain structures are not identical.

The $(7,4,2)$ biplane of Example 1.2 will have $v-k=7-4=3$ chains comprising each of its $b=7$ possible chain structures. Each chain is necessarily a 4-cycle. There is only one possible chain associated with $\{1,2\}$ and $\{1,3\}$, namely $(1,2,4,3)$. (Of course $(1,2,4,3)$, $(2,4,3,1)$, $(3,4,2,1)$, etc., all represent the same chain.) The same is true for the other two pairs of

2-subsets. Thus the only possible distinct chains are $(1,2,4,3)$, $(1,2,3,4)$, and $(1,3,2,4)$, and they do determine the biplane of Example 1.2. Each of the seven chain structures must therefore be the same. This argument actually establishes the uniqueness of the biplane of order 2.

For biplanes of order 4, there are two possible chain types, namely, (6)-chains and (3-3)-chains. The number of chains comprising a given chain structure is $v-k=16-6=10$. Consider the chain associated with the pair of 2-subsets $\{1,2\}$ and $\{1,3\}$. There are one (3-3)-chain and six (6)-chains from which to choose, and the same is true for each of the other nine pairs of 2-subsets. The reader may verify, quite simply, that the choice of the (3-3)-chain in each instance is a correct choice (each pair of points is contained in precisely two blocks). This yields the biplane whose incidence matrix has a 2-rank of 6, which we have previously called $\mathcal{B}_6$. Each of the sixteen chain structures for $\mathcal{B}_6$ consists of these ten (3-3)-chains. For $\mathcal{B}_7$, each chain structure consists of four (6)-chains and six (3-3)-chains, and for $\mathcal{B}_8$, each structure consists of six and four, respectively.

As one further example of this λ-chain notion, we will examine one of the biplanes of order 7. The following group-theoretic description can be found in [**15**]. Let $\mathcal{P}$ consist of the 1-dimensional subspaces of $PG(1,8)$ (see Section 2). Note that $PGL_2(8)$, the projective general linear group of 2×2 matrices with entries from the field with 8 elements, acts sharply triply transitively on $\mathcal{P}$ (see [**10**, Corollary 2.6.5]); that is, given any two ordered triples of elements of $\mathcal{P}$, there exists a unique element of $PGL_2(8)$ which sends the first ordered triple to the second. Let the set of points of the biplane consist of the nine elements of $\mathcal{P}$ and the $\binom{9}{3}/3$ subgroups of order 3 of $PGL_2(8)$. Each of these 28 subgroups is generated by an element of order 3 with no fixed points, and the orbit structure is that of a chain with all 3-cycles. One block, indexed by $\mathcal{P}$, consists of the 9 points of $\mathcal{P}$. Each of the remaining 36 blocks is indexed by a 2-subset and such a block consists of the 2-subset and the 7 subgroup-of-order-3 points which contain the 2-subset in one of its orbits. Originally discovered by Hussain [**34**], this biplane's chain structures, and those of its nonisomorphic dual (presented below on the right), are:

1	with 28	(3-3-3)-chains	9	with 28	(6-3)-chains
36	with 21	(9)-chains	28	with 27	(9)-chains
	7	(6-3)-chains		1	(3-3-3)-chain.

Although the 7-rank does not aid in distinguishing the four biplanes of order 7, the λ-chain idea does since the collection of chain structures for each biplane is different [**49**].

Additional information on the chain structures of the known biplanes can be found in [**48**], [**49**], and [**50**].

Let $D=\{g_1,g_2,\ldots,g_k\}$ be a k-subset of a group G with v elements. If for each nonidentity element d of G there exist precisely λ ordered pairs (g_i,g_j) of elements of D such that $g_ig_j^{-1}=d$, then we call D a (v,k,λ) *group difference set*. If $G=\mathbb{Z}_v$, then D is simply called a (v,k,λ) *difference set*. (See [**13**] and [**24**].)

PROPOSITION 3.1. *If D is a (v,k,λ) group difference set composed of elements of the group G, then the sets $Dg=\{g_1g,g_2g,\ldots,g_kg\}$, for all $g\in G$, form the blocks of a projective (v,k,λ) design. (If $G=\mathbb{Z}_v$, such a design is called a cyclic design.)*

Proof. Let a and b be any two distinct elements of G. We wish to exhibit precisely λ blocks which contain these two elements. Let $d=ab^{-1}$. There exist precisely λ ordered pairs (g_i,g_j) such that $g_ig_j^{-1}=d=ab^{-1}$. Thus $g_i^{-1}a=g_j^{-1}b$ are the λ g's such that Dg contains both a and b. Since v blocks are constructed in this manner, we have a projective design.

Every Desarguesian plane is, in fact, a cyclic plane, and no other cyclic planes are known. For the method of constructing these difference sets see [**51**] and Chapter 11 of [**24**].

If one relabels "7" as "0" in Example 1.1, one obtains the construction of the plane of order 2

from the difference set $D=\{1,2,4\}$, which consists of the quadratic residues (mod 7) (see [**24**, p. 141]). Similarly, the biplane of order 2 of Example 1.2 is constructed from the difference set $D=\{1,2,3,6\}$. The unique biplane of order 3 can be constructed via the $(11,5,2)$ difference set $D=\{1,3,4,5,9\}$, which consists of the quadratic residues (mod 11). Also, another one of the biplanes of order 7, discovered independently by Bose [**12**] and Fisher [**19**], can be constructed from the difference set which consists of the biquadratic residues (mod 37).

Each of the three biplanes of order 4 can be constructed via $(16,6,2)$ group difference sets. (See [**35**] and [**7**] for a detailed discussion.) For example, the group difference set

$$D=\{(0,0),(0,1),(0,2),(0,5),(1,0),(1,6)\}$$

in $\mathbb{Z}_2\times\mathbb{Z}_8$ [**54**] yields the biplane of 2-rank 6, which we have called $\mathfrak{B}_6$.

We now turn to the final topic of this section, which is the notion of *oval*. Although of great interest in the theory of planes [**18**], the notion of oval has only recently been generalized by Assmus and van Lint [**8**] to arbitrary projective designs. We present here but a few of their results.

Let $(S,\mathfrak{D})$ be a projective (v,k,λ) design. An *arc* $\mathcal{C}$ is a subset of S with the property that no three points of $\mathcal{C}$ lie on a block. That is, for each $B\in\mathfrak{D}, |B\cap\mathcal{C}|=0$, 1, or 2, and B is called an *exterior*, *tangent*, or *secant* block, respectively.

THEOREM 3.2. *If $(S,\mathfrak{D})$ is a projective (v,k,λ) design with $k>2$ and the design is of either odd order or even order with $k\not\equiv 0 \pmod\lambda$, then $|\mathcal{C}|\leqslant(1/\lambda)(k+\lambda-1)$. If it is of even order with $k\equiv 0 (\operatorname{mod}\lambda)$, then $|\mathcal{C}|\leqslant(1/\lambda)(k+\lambda)$.*

An arc, $\mathcal{C}$, of the projective design $(S,\mathfrak{D})$ is called an *oval*, $\mathcal{O}$, whenever it achieves the prescribed bound. We denote the collection of all ovals of the design by *Oval* $(\mathfrak{D})$.

COROLLARY. *Let $n=k-\lambda$ denote, as usual, the order of the projective design.*

(1) *If $\mathcal{O}$ is an oval of a plane of order n, then $|\mathcal{O}|=n+1$ if n is odd and $n+2$ if n is even.*

(2) *If $\mathcal{O}$ is an oval of a biplane of order n, then $|\mathcal{O}|=\frac{1}{2}(n+3)$ if n is odd and $\frac{1}{2}(n+4)$ if n is even.*

PROPOSITION 3.3. *If $\mathcal{O}$ is an oval in a projective design of even order with $k\equiv 0(\operatorname{mod}\lambda)$, then $\mathcal{O}$ has $(1/2\lambda)k(k+\lambda)$ secants, $(1/2\lambda)(k-2)(k-\lambda)$ exterior blocks, and no tangents. If $\mathcal{O}$ is an oval in a projective design of odd order, then $\mathcal{O}$ necessarily has $(1/\lambda)(k+\lambda-1)$ tangents, $(1/2\lambda)(k+\lambda-1)(k-1)$ secants, and $(1/2\lambda)(k-\lambda-1)(k-1)$ exterior blocks. Moreover, in the odd order case, through each point of $\mathcal{O}$ there passes exactly one tangent, through each point not on $\mathcal{O}$ there pass either two tangents (exterior point) or none (interior point), and thus the tangents form an oval, $\mathcal{O}^d$, in the dual design.*

Let us now examine the oval structure of some planes and biplanes. If there is a unique biplane of order n, we shall denote it by $\mathfrak{B}(n)$.

Let $\mathcal{O}$ be an oval of PG(2,2), the plane of order 2; then $\mathcal{O}$ is a 4-subset that does not contain a line and so

$$|\mathrm{Oval}(\mathrm{PG}(2,2))|=\binom{7}{4}-7\cdot 4=7;$$

that is, Oval (PG(2,2)) $=\mathfrak{B}(2)$ (see Examples 1.1 and 1.2). Also, an oval $\mathcal{O}$ of $\mathfrak{B}(2)$ is a 3-subset not contained in any block of $\mathfrak{B}(2)$, and so

$$|\mathrm{Oval}(\mathfrak{B}(2))|=\binom{7}{3}-7\cdot 4=7;$$

that is, Oval $(\mathfrak{B}(2))=$ PG(2,2).

The oval situation for all known odd order biplanes is as follows.

$n=1$: $\mathfrak{B}(1)$ is the complete $(4,3,2)$ biplane whose blocks are all 3-subsets of a 4-set S. Since

an oval contains two elements of S, Oval ($\mathfrak{B}(1)$) consists of the $\binom{4}{2}=6$ 2-subsets of S.

$n=3$: Here $|\mathfrak{O}|=3$, and so an oval is a 3-subset which is not contained in a block. Thus

$$|\mathrm{Oval}(\mathfrak{B}(3))|=\binom{11}{3}-11\cdot\binom{5}{3}=55.$$

The Lehigh University CDC 6400 computer was employed in establishing the existence of precisely four biplanes of order 7 (λ-chains were employed in the exhaustive search), constructing the last biplane of order 7 and one of the four known biplanes of order 9, and examining the oval structure for the biplanes of orders 7, 9, and 11 (see **[49]** and **[50]**). (When both the size of an oval and k exceed 3, the situation is of course more complicated than we saw above.) A group-theoretical description of the oldest biplane of order 9 is in **[26]** and the last two were discovered by R. H. F. Denniston.

$n=7$: The difference set biplane has no ovals. Each of the other three has 63 ovals. In particular, the biplane constructed via $\mathrm{PGL}_2(8)$ has the block indexed by $\mathrm{PG}(1,8)$ tangent to each of its ovals, and this is not true of any other block. This last result is crucial to one of the constructions we present in Section 4.

$n=9$: The 3-ranks of the four known biplanes are 20, 22, 24, and 26, and the number of ovals present are 336, 120, 64, and 48, respectively. The largest number of ovals which have a common tangent is 36. It is interesting to note that, for these four biplanes, as the 3-rank increases, the number of ovals decreases. Also, the chain structures for these biplanes admit cycles of longer length as the 3-rank increases. Perhaps these statements can be shown to be true in general. We shall certainly have occasion to refer to these conjectures again in the conclusion to Section 4. (See **[50]** for further information on these biplanes and their automorphism groups.)

$n=11$: The two known biplanes, constructed by Aschbacher **[2]**, are duals of each other, and hence, by Proposition 3.3, admit the same number of ovals, which is 77. Here the largest number of ovals which share a common tangent is 55.

4. Planes from Biplanes. We first present four methods of producing self-orthogonal codes via the incidence matrices of projective designs. (See **[6]**.)

THEOREM 4.1. *Let M be the incidence matrix of a projective (v,k,λ) design.*

(1) *If $k\equiv\lambda\equiv 0 \pmod p$, then A, the rowspace of M over F_p, is a self-orthogonal code.*

(2) *If $p\,|\,n$, where $n=k-\lambda$, but $p\nmid k$, let G be the $v\times(v+1)$ matrix whose first column consists of $\sqrt{-k}$'s and whose last v columns are those of M, and let $F=F_p$ if $-k$ is a quadratic residue* (mod p) *and F_{p^2} otherwise. Then A, the rowspace of G over F, is a self-orthogonal code. Moreover, if $p^2\nmid n$, then $A=A^{\perp}$ and so A is a $(v+1,\frac{1}{2}(v+1))$ self-dual code over F.*

(3) *If $k+1\equiv\lambda\equiv 0 \pmod p$, let G be the $v\times 2v$ matrix whose first v columns constitute the identity matrix and whose last v columns are the columns of M. Then A, the rowspace of G over F_p, is a $(2v,v)$ self-dual code.*

(4) *If $p=2$, λ is odd, and k is even, let G be the $(v+1)\times(2v+2)$ matrix whose first $v+1$ columns constitute the identity matrix, whose $(v+2)$-column consists of a 0 in the first row and 1's elsewhere, and whose last v columns are those of M bordered above by a row of 1's. Then A, the rowspace of G over F_2, is a $(2v+2,v+1)$ self-dual code.*

Proof. The proofs of (1), (3), and (4) are straightforward and are left to the reader. (In establishing (4), simply note that $k(k-1)=\lambda(v-1)$ implies that v is odd.)

We present a proof of part (2). A is clearly self-orthogonal. Since $A\subseteq A^{\perp}$ and $\dim_F(A)+\dim_F(A^{\perp})=v+1$, $\mathrm{Rk}_F(G)\leqslant\frac{1}{2}(v+1)$. Now $\mathrm{Rk}_F(G)=\mathrm{Rk}_F(M)$ and, by Proposition 2.2, $\mathrm{Rk}_F(M)\geqslant v-\frac{1}{2}(v-1)=\frac{1}{2}(v+1)$ when $p^2\nmid n$. Thus $\mathrm{Rk}_F(G)=\frac{1}{2}(v+1)$ and so A is a self-dual code.

Any biplane of even order with $p=2$ is an example for Method (1).

For (2), take any plane and a prime that divides the order. For any plane of order p, we again obtain $\mathrm{Rk}_p(M)=\frac{1}{2}(v+1)$ (see Proposition 2.3). It is interesting to note that, should a plane of order 10 exist, this theorem guarantees the 2-rank of its incidence matrix to be $\frac{1}{2}(111+1)=56$.

By taking p to be 2, any biplane of odd order is an example for Method (3); and for (4), take any plane of odd order.

See [6] for additional examples of this theorem which yield perfect codes. Also in [6], Assmus, Mezzaroba, and Salwach present the following technique for producing planes from biplanes via Method (3).

THEOREM 4.2. *Let M be the incidence matrix of a biplane of odd order with parameters $(v,k,2)$, and let I be a $v \times v$ identity matrix. Let G be the $v \times 2v$ matrix with the v columns of I preceding those of M and let G' be the $v \times 2v$ matrix with the columns of M^t, the incidence matrix of the dual biplane, preceding those of I. Then over F_2, $\mathrm{SP}(G) = \mathrm{SP}(G')$ and this subspace of F_2^{2v}, call it A, is a self-dual code with minimum-weight $d(A) = k+1$. Moreover, the minimum-weight vectors which are neither the rows of G nor the rows of G' have precisely half of their 1's in the first v coordinates.*

Proof. $G=(I|M)$ and $G'=(M^t|I)$. Since $\lambda=2$ and k is odd, $MM^t = I \pmod 2$ and thus G' may be obtained from G via the standard row operations involved in computing the inverse of a matrix. Hence $\mathrm{SP}(G) = \mathrm{SP}(G')$. A is self-dual by virtue of Method (3) of Theorem 4.1. Since each row of G or G' has weight $k+1, d(A) \leqslant k+1$. In order to demonstrate the remaining assertions, it suffices to establish that the (mod 2) sum of s distinct rows of G, where $2 \leqslant s \leqslant \frac{1}{2}(k-1)$, is a vector with weight greater than $k+1$, since the same argument also establishes the result for G' and any vector providing a counterexample would necessarily have less than $\frac{1}{2}(k+1)$ 1's in either the first or second set of v coordinates.

By induction on s, the number of rows, we show the (mod 2) sum of s rows has weight at least $s(k+3-2s)$. The weight of one row is $1\cdot(k+1)$ and the weight of the sum of two rows is $2\cdot(k-1)$, thereby meeting the bound. Assuming the result for s, the sum of $s+1$ distinct rows has weight at least

$$s(k+3-2s)+(k+1)-4s=(s+1)[k+3-2(s+1)],$$

since any row will meet the sum of s rows in at most $2s$ coordinates inasmuch as half of the row is a block of a biplane.

Since $s(k+3-2s)$ is a parabola which is concave downward, when considered as a function of s, and for $s=\frac{1}{2}(k+1)$ we obtain the value of $k+1$ the same as for $s=1$, the sum of $s\leqslant\frac{1}{2}(k+1)$ rows produces a vector of weight at least $k+1$.

Note that for a biplane, $v=\frac{1}{2}(k^2-k+2)$, and so

$$2v=k^2-k+2=(k-1)^2+(k-1)+1+1,$$

one more than the number of points in a plane of order $k-1$. Thus we attempt to extract the even order plane from the $(2v,v)$ code obtained from the odd order biplane by selecting all of the weight-$(k+1)$ vectors with a 1 at a given coordinate, call this collection $\mathfrak{L}$, and then determine whether the collection of weight-k vectors obtained by simply ignoring the given coordinate contain all of the lines of a plane of order $k-1$ on the remaining k^2-k+1 coordinates. We call this process "contracting on a point." Note: The collection of all weight-$(k+1)$ vectors need not, in general, form a design.

PROPOSITION 4.3. *If the collection $\mathfrak{L}$ yields all of the lines of a plane of order $k-1$, then these lines "constitute" all of $\mathfrak{L}$. Moreover, the supports of all other weight-$(k+1)$ vectors of A are ovals of the plane of order $k-1$ just obtained.*

Proof. Let P be the point on which we contract. Let N be the $(k^2-k+1)\times(k^2-k+2)$ matrix whose rows are the weight-$(k+1)$ vectors of $\mathfrak{L}$ which yield the lines of the plane.

$A \subseteq (\mathrm{SP}(N))^{\perp}$. Say there exists $w \in \mathfrak{L}$ such that $m = \mathrm{supp}(w) - \{P\}$ is not a line. $|m| = k$. Since $w \in (\mathrm{SP}(N))^{\perp}$ and w meets each row of N in P, m must meet each line of the plane in at least one point. Let $Q, R \in m$. Let l be the unique line containing Q and R. Now, there exists $S \in l - m$ and there exist $k-1$ lines other than l through S, each meeting m. Hence $|m| \geqslant (k-1)+2$, a contradiction.

Let $w \in A$ be such that $w \notin \mathfrak{L}$ and $\mathrm{wgt}(w) = k+1$. Let $Q \in \mathrm{supp}(w)$. Let l be any of the k lines through Q. Since $w \in (\mathrm{SP}(N))^{\perp}$, $|l \cap \mathrm{supp}(w)|$ is even, and so each l meets $\mathrm{supp}(w)$ in at least one point other than Q. Since $|\mathrm{supp}(w)| = k+1$, it must be exactly once more (two lines of a plane meet in precisely one point). Therefore $\mathrm{supp}(w)$ is an oval of the plane.

Note that the preceding paragraph is the essential ingredient in proving that $d(\mathrm{SP}(M)^{\perp}) = n + 2$ and that the weight-$(n+2)$ vectors of $(\mathrm{SP}(M))^{\perp}$ are precisely the ovals of the plane of even order n with incidence matrix M.

Since $G = (I|M)$ and $G' = (M^t|I)$ both span A (see Theorem 4.2), we index the first v columns of G by the blocks of the biplane and the last v, as usual, by the points. Assmus and van Lint **[8]** have characterized the minimum-weight vectors of A which are neither rows of G nor of G' in terms of biplane ovals as follows:

PROPOSITION 4.4. *Let $\mathcal{O}$ be an oval of a biplane of odd order, then $\mathcal{O}^d \cup \mathcal{O}$ is the support of a minimum-weight vector of A (of Theorem 4.2). Conversely, if w is a minimum-weight vector of A which has half its 1's in the first v coordinates, then $\mathrm{supp}(w) = \mathcal{O}^d \cup \mathcal{O}$, where $\mathcal{O}$ is an oval of the biplane.*

Proof. Let $\mathcal{O}$ be an oval of the biplane of odd order, then $\mathcal{O}^d$ consists of its tangents. Form a vector $w \in F_2^{2v}$ by placing 1's in the $\frac{1}{2}(k+1)$ coordinates indexed by the tangents and the $\frac{1}{2}(k+1)$ coordinates indexed by the points of $\mathcal{O}$. We now show that $w \in A$. Since through each point of $\mathcal{O}$ there passes exactly one tangent and through each point not on $\mathcal{O}$ there pass two tangents or none (see Proposition 3.3), the (mod 2) sum of the rows of G indexed by the tangents will yield a vector whose support is $\mathcal{O}^d \cup \mathcal{O}$. Therefore $\mathcal{O}^d \cup \mathcal{O}$ gives rise to a minimum-weight vector of A which is of the desired form.

Conversely, let w be a minimum-weight vector of A of the correct form. Let $\mathcal{O}$ be the support of the last v coordinates of w. We now show that $\mathcal{O}$ is a biplane oval. Let $P \in \mathcal{O}$. There exist $\frac{1}{2}(k-1)$ other points in $\mathcal{O}$. Since through P and each other point of $\mathcal{O}$ there pass two blocks of the biplane, at most $k-1$ blocks through P are accounted for. Hence there exists at least one block B such that $B \cap \mathcal{O} = \{P\}$. Since $A = A^{\perp}$, there is a 1 in the coordinate of w indexed by B. Thus each of the $\frac{1}{2}(k+1)$ 1's in the first v coordinates of w are accounted for by a single such B for each $P \in \mathcal{O}$. Hence each such B is unique, and so each of the remaining $k-1$ blocks through P must meet $\mathcal{O}$ in precisely one more point. Thus $\mathcal{O}$ is an oval of the biplane and the support of the first v coordinates of w forms $\mathcal{O}^d$.

We conclude by examining the odd order biplanes in light of this construction. See **[6]**, **[8]**, **[42]**, **[48]**, and **[50]** for more detailed information on these intriguing examples.

$n=1$: The $(4,3,2)$ biplane yields, via Theorem 4.2, the doubly even self-dual extended binary $(8,4)$ Hamming code whose 14 weight-4 vectors form the 3-$(8,4,1)$ design of Example 1.3 and we show this as follows: Two distinct weight-4 vectors meet in either zero or two 1's and so each 3-subset of the $\binom{8}{3} = 56$ possible is contained in the support of at most one weight-4. But each of the 14 weight-4's has $\binom{4}{3} = 4$ 3-subsets yielding a total of 56, so each 3-subset is contained in the support of precisely one weight-4. This 3-$(8,4,1)$ design is the unique extension of the plane of order 2, the 2-$(7,3,1)$ design of Example 1.1. Contracting on any point yields the plane of order 2. Also, as we saw in Section 3, $\mathfrak{B}(1)$ has 6 ovals. These 6 (by virtue of Proposition 4.4) together with the 4 rows of G and the 4 rows of G' yield the 14 weight-4 vectors.

$n=3$: The $(11,5,2)$ biplane yields a $(22,11)$ self-dual code A with $d(A)=6$. We show that the collection of all weight-6 vectors forms a 3-$(22,6,1)$ design, and so will be the extension of the plane of order 4, a 2-$(21,5,1)$ design. Let b_i and b_j be any two distinct vectors of weight 6 in A. Now

$$|\operatorname{supp}(b_i)\cap\operatorname{supp}(b_j)|\leqslant 2$$

since if it were at least 3, then, because $A=A^{\perp}$, it would be 4, but then $\operatorname{wgt}(b_i+b_j)=4<6=d(A)$, a contradiction. Thus each 3-subset is in the support of at most one weight-6 vector. Since the rows of G and G' yield 22 weight-6 vectors, and $\mathfrak{B}(3)$ admits 55 ovals, thereby yielding, via Proposition 4.4, 55 weight-6 vectors which "split 3-3", and $77\cdot\binom{6}{3}=\binom{22}{3}$, we see that A contains 77 weight-6 vectors and that each 3-subset is in the support of precisely one weight-6 vector.

$n=7$: All four $(37,9,2)$ biplanes yield $(74,37)$ self-dual codes. The difference set biplane cannot yield the plane of order 8, since difference set projective designs have transitive automorphism groups and are isomorphic to their duals. Hence the $(74,37)$ code of this biplane has a transitive automorphism group and so, if a plane was obtained by contracting at one point, it would be obtained by contracting at any point, a contradiction, since only the planes of orders 2 and 4 (and possibly 10, if it exists), have extensions (see Theorem 1.11).

Only the biplane gotten via $PGL_2(8)$ yields the plane of order 8, and only by contracting on the first coordinate (indexed by $\mathcal{P}=PG(1,8)$). Since $73=1+9+63$, we yet need 63 weight-10 vectors which split 5-5. Since $PGL_2(8)$ acts sharply triply transitively on $\mathcal{P}$, there exist $9\cdot\binom{8}{2}/4=63$ subgroups of order 2, each of whose orbits consist of a fixed point and four 2-subsets. The (mod 2) sum of the rows indexed by $\mathcal{P}$ and the four 2-subsets yield a 5-5 weight-10 vector with a 1 in the first coordinate. These 63 weight-10 vectors can, of course, be constructed as well from the 63 ovals of the biplane, each of which has as a tangent the block indexed by $\mathcal{P}$. These 73 vectors do yield the plane of order 8 on coordinates 2 through 74 (see **[48]**). It is interesting to note that we contracted on the unique indexing block that yielded all (3-3-3)-chains. Of course, the chain structure for $\mathfrak{B}$ (1) consists of a (3)-chain, and for $\mathfrak{B}$ (3) it consists of (5)-chains.

$n=9$: One of the four known $(56,11,2)$ biplane (in fact, the one presented in **[26]**) can be obtained via the $56=77-21$ unused weight-6 vectors in the case of $n=3$ when contracting to produce the plane of order 4; that is, the 56 weight-6 vectors with a 0 at a given coordinate. Let b_i and b_j be any two such vectors. Define M by $m_{ij}=0$ if $b_i\cdot b_j=2$ and 1 otherwise. M is the incidence matrix of the biplane of order 9 whose 3-rank is 20.

Unfortunately, none of the four known biplanes of order 9 yield the plane of order 10, since none admit enough ovals. In fact, as we previously observed in Section 3, the number of ovals of a biplane of order 9 seems to decrease as the 3-rank increases and as the length of the λ-chain cycles increase. The same is also true for the biplanes of order 4 **[7]**, **[8]**. An exhaustive computer search demonstrated the biplanes of 3-rank 20 and 22 to be the only biplanes of order 9 with at least one chain structure consisting entirely of (4-4-3)-chains (see **[50]**). Thus hopes of producing a plane of order 10 via these techniques do not seem very bright.

$n=11$: The two known $(79,13,2)$ biplanes are duals of each other and do not yield the plane of order 12. Again, the biplanes do not admit enough ovals. A computer search has shown that there does not exist a biplane of order 11 with at least one chain structure consisting entirely of (4-3-3-3)-chains, the seemingly most interesting case.

References

1. W. O. Alltop, Extending t-designs, J. Combin. Theory Ser. A, 18 (1975) 177–186.

2. M. Aschbacher, On collineation groups of symmetric block designs, J. Combin. Theory, 11 (1971) 272–281.

3. E. F. Assmus, Jr., and D. P. Maher, Nonexistence proofs for projective designs, this MONTHLY, 85 (1978) 110–112.

4. E. F. Assmus, Jr., and H. F. Mattson, Jr., Coding and combinatorics, SIAM Rev., 16 (1974) 349–388.

5. E. F. Assmus, Jr., H. F. Mattson, Jr., and M. Guza, Self-orthogonal Steiner systems and projective planes, Math. Z., 138 (1974) 89–96.

6. E. F. Assmus, Jr., J. A. Mezzaroba, and C. J. Salwach, Planes and biplanes, in M. Aigner (ed.), Higher Combinatorics, D. Reidel, Dordrecht, Holland, 1977, pp. 205–212.

7. E. F. Assmus, Jr., and C. J. Salwach, The (16,6,2) designs, Internat. J. Math. Math. Sci., 2 (1979) 261–281.

8. E. F. Assmus, Jr., and J. H. van Lint, Ovals in projective designs, J. Combin. Theory Ser. A, 27 (1979) 307–324.

9. E. Berlekamp, Algebraic Coding Theory, McGraw-Hill, New York, 1968.

10. N. Biggs, Finite Groups of Automorphisms, London Math. Soc. Lecture Note Series, no. 6, Cambridge University Press, 1971.

11. J. W. Blattner, Projective Plane Geometry, Holden-Day, San Francisco, 1968.

12. R. C. Bose, On the construction of balanced incomplete block designs, Ann. Eugenics, 9 (1939) 353–399.

13. R. H. Bruck, Difference sets in a finite group, Trans. Amer. Math. Soc., 78 (1955) 464–481.

14. R. H. Bruck and H. J. Ryser, The nonexistence of certain finite projective planes, Canad. J. Math., 1 (1949) 88–93.

15. P. J. Cameron, Biplanes, Math. Z., 131 (1973) 85–101.

16. P. J. Cameron and J. H. van Lint, Graph Theory, Coding Theory and Block Designs, London Math. Soc. Lecture Note Series, no. 19, Cambridge University Press, 1975.

17. S. Chowla and H. J. Ryser, Combinatorial problems, Canad. J. Math., 2 (1950) 93–99.

18. P. Dembowski, Finite Geometries, Springer-Verlag, Berlin-Heidelburg-New York, 1968.

19. R. A. Fisher, An examination of the different possible solutions of a problem in incomplete blocks, Ann. Eugenics, 10 (1940) 52–75.

20. J.-M. Goethals and P. Delsarte, On a class of majority logic decodable cyclic codes, IEEE Trans. Inform. Theory, 14 (1968) 182–188.

21. R. L. Graham and J. MacWilliams, On the number of information symbols in difference-set cyclic codes, Bell System Tech. J., 45 (1966) 1057–1070.

22. M. Hall, Jr., Finite projective planes, in Contributions to Geometry, Slaught Memorial Papers, no. 4, this MONTHLY, 62 (1955) 18–24.

23. ______, The Theory of Groups, Macmillan, New York, 1959.

24. ______, Combinatorial Theory, Blaisdell, Waltham-Toronto-London, 1967.

25. ______, Construction of block designs, in J. N. Srivastava, et al. (eds.), A Survey of Combinatorial Theory, North-Holland, London, 1973.

26. M. Hall, Jr., R. Lane, and D. Wales, Designs derived from permutation groups, J. Combin. Theory, 8 (1970) 12–22.

27. M. Hall, Jr., J. D. Swift, and R. B. Killgrove, On projective planes of order 9, Math. Comp., 13 (1959) 233–246.

28. M. Hall, Jr., J. D. Swift, and R. J. Walker, Uniqueness of the projective plane of order eight, Math. Tables and Aids to Computation, 10 (1956) 186–194.

29. N. Hamada, On the p-rank of the incidence matrix of a balanced or partially balanced incomplete block design and its applications to error-correcting codes, Hiroshima Math. J., 3 (1973) 154–226.

30. N. Hamada and H. Ohmori, On the BIB design having the minimum p-rank, J. Combin. Theory Ser. A, 18 (1975) 131–140.

31. D. R. Hughes, On t-designs and groups, Amer. J. Math., 87 (1965) 761–778.

32. D. R. Hughes and F. C. Piper, Projective Planes, Springer-Verlag, Berlin-Heidelberg-New York, 1973.

33. Q. M. Hussain, On the totality of the solutions for the symmetrical incomplete block designs $\lambda=2$, $k=5$ or 6, Sankhyā, 7 (1945) 204–208.

34. ______, Symmetrical incomplete block designs with $\lambda = 2$, $k = 8$ or 9, Bull. Calcutta Math. Soc., 37 (1945) 115–123.

35. R. E. Kibler, A summary of noncyclic difference sets, $k < 20$, J. Combin. Theory Ser. A, 25 (1978) 62–67.

36. E. Kleinfeld, Techniques for enumerating Veblen-Wedderburn systems, J. Assoc. Comput. Mach., 7 (1960) 330–337.

37. P. Lorimer, A projective plane of order 16, J. Combin. Theory, 16 (1974) 334–347.

38. F. J. MacWilliams and H. B. Mann, On the p-rank of the design matrix of a difference set, Inform. and Control, 12 (1968) 474–488.

39. F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, North-Holland, Amsterdam, 1977.

40. F. J. MacWilliams, N. J. A. Sloane, and J. G. Thompson, On the existence of a projective plane of order 10, J. Combin. Theory, 14 (1973) 66–78.

41. N. S. Mendelsohn, Intersection numbers of t-designs, in Studies in Pure Mathematics, Academic Press, New York, 1971, pp. 145–150.

42. J. A. Mezzaroba, Weights and Automorphisms of Linear Codes, Ph.D. dissertation, Lehigh University, Bethlehem, Pa., 1975.

43. W. W. Peterson and E. J. Weldon, Jr., Error-Correcting Codes, 2nd ed., MIT Press, Cambridge, 1972.

44. H. J. Ryser, Geometries and incidence matrices, in Contributions to Geometry, Slaught Memorial Papers, no. 4, this MONTHLY, 62 (1955) 25–31.

45. ______, Combinatorial Mathematics, Carus Mathematical Monographs, no. 14, Math. Assoc. of Amer., 1963.

46. H. E. Sachar, Error-Correcting Codes Associated with Finite Projective Planes, Ph.D. dissertation, Lehigh University, Bethlehem, Pa., 1973.

47. ______, The F_p span of the incidence matrix of a finite projective plane, Geom. Dedicata, 8 (1979) 407–415.

48. C. J. Salwach, Biplanes and Projective Planes, Ph.D. dissertation, Lehigh University, Bethlehem, Pa., 1976.

49. C. J. Salwach and J. A. Mezzaroba, The four biplanes with $k=9$, J. Combin. Theory Ser. A, 24 (1978) 141–145.

50. ______, The four known biplanes with $k=11$, Inter. J. Math. Math. Sci., 2 (1979) 251–260.

51. J. Singer, A theorem in finite projective geometry and some applications to number theory, Trans. Amer. Math. Soc., 43 (1938) 377–385.

52. K. J. C. Smith, On the p-rank of the incidence matrix of points and hyperplanes in a finite projective geometry, J. Combin. Theory, 7 (1969) 122–129.

53. F. W. Stevenson, Projective Planes, W. H. Freeman, San Francisco, 1972.

54. R. J. Turyn, Character sums and difference sets, Pacific J. Math., 15 (1965) 319–346.

55. J. H. van Lint, Coding Theory, Lecture Notes in Mathematics, no. 201, Springer-Verlag, Berlin-Heidelberg-New York, 1971.

56. E. G. Whitehead, Jr., Difference sets and sum-free sets in groups of order 16, Discrete Math., 13 (1975) 399–407.

57. F. Yates, Incomplete randomized blocks, Ann. Eugenics, 7 (1936) 121–140.

SOME HISTORICAL REMARKS CONCERNING DEGREE THEORY

HANS WILLI SIEGBERG

Institut für Angewandte Mathematik der Universität Bonn, Wegelerstr. 6, Federal Republic of Germany

Summary. It is well known that the notion of topological degree, $\deg(F)$, of a continuous mapping $F: M \to N$, where M and N are connected, oriented, compact n-manifolds (with triangulation), can be traced back to L. E. J. Brouwer [Brouwer, 1911, 1912]. A related notion is that of the winding number, $w(F,0)$, of a continuous mapping $F: M \to \mathbb{R}^{n+1}-\{0\}$ $(w(F,0)=\deg(F/\|F\|))$. In the differentiable case this concept was known as the "Kronecker characteristic" or the "Kronecker integral" of the mapping F even before Brouwer's work. The aim of this note is to examine the historical roots of the Kronecker characteristic and some of its applications.

1. Introduction. A familiar representation of the **Kronecker characteristic** is given by the following integral; see, e.g., [Hadamard, 1910] and [Alexandroff and Hopf, 1935]:

Let $F: M \to \mathbb{R}^{n+1}-\{0\}$ be a smooth map, and let M be a smooth, compact, and oriented

The author received a diploma in Mathematics and Information Theory in 1978 under H. O. Peitgen at the University of Bonn; his work is currently supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 72, at the university. From September 1979 to February 1980 he was a Research Scholar at the University of Utah. His current interests are in fixed-point algorithms and the numerical computation of topological degree.—*Editors*

n-manifold:

$$K(F):=(\operatorname{vol} S^n)^{-1}\int_M \|F(x)\|^{-(n+1)}\det\left(F(x),\frac{\partial F}{\partial x_1}(x),\ldots,\frac{\partial F}{\partial x_n}(x)\right)dx$$

$(x=(x_1,\ldots,x_n)$ local coordinates of $M)$.

In terms of differential forms we have

$$K(F)=(\operatorname{vol} S^n)^{-1}\int_M (r\circ F)^*\sigma,$$

where $r:\mathbb{R}^{n+1}-\{0\}\to S^n$ is the map $r(x):=x/\|x\|$, and

$$\sigma:=\sum_{i=1}^{n+1}(-1)^{i+1}x_i\,dx_1\wedge\cdots\wedge\overset{\wedge}{dx_i}\wedge\cdots\wedge dx_{n+1}|_{S^n}$$

is the volume form of S^n.

$K(F)$ is called the **Kronecker integral** of the mapping F, and it can be shown that $K(F)=w(F,0)=\deg(r\circ F)$.

Actually Kronecker introduced his characteristic in a different way [Kronecker, 1869a], and his motivation seems to have been highly influenced by an attempt to generalize Sturm's theorem.

In order to see these relations we will first describe Kronecker's definition in modern language.

2. The Kronecker Characteristic. The "original" Kronecker characteristic is defined for certain mappings $F:\mathbb{R}^{n+1}\to\mathbb{R}^{n+2}-\{0\}$ rather than for mappings $M\to\mathbb{R}^{n+1}-\{0\}$.

DEFINITION. A smooth map

$$F:=(F_0,\ldots,F_{n+1}),\quad F:\mathbb{R}^{n+1}\to\mathbb{R}^{n+2}-\{0\}\qquad(n\geqslant 1),$$

is called a **regular function system** if the following properties hold:

(i) $0\in\operatorname{int}F_i(\mathbb{R}^{n+1})$ for $0\leqslant i\leqslant n+1$.
(ii) $F_i^{-1}(0)$ is compact for $0\leqslant i\leqslant n+1$.
(iii) $0\in\mathbb{R}$ is a regular value for F_i for $0\leqslant i\leqslant n+1$, i.e., $F_i(x)=0\Rightarrow dF_i(x)$ is of full rank.
(iv) $0\in\mathbb{R}^{n+1}$ is a regular value for the mapping

$$(F_0,\ldots,\hat{F_i},\ldots,F_{n+1}):\mathbb{R}^{n+1}\to\mathbb{R}^{n+1}$$

for $0\leqslant i\leqslant n+1$.
(v) $0\in\mathbb{R}^n$ is a regular value for the mapping

$$(F_0,\ldots,\hat{F_i},\ldots,\hat{F_j},\ldots,F_{n+1}):\mathbb{R}^{n+1}\to\mathbb{R}^n$$

for $0\leqslant i,j\leqslant n+1$, and $i\neq j$.

By virtue of these properties it follows that (1) the set

$$F(\{i,j\}):=\left(F_0,\ldots,\hat{F_i},\ldots,\hat{F_j},\ldots,F_{n+1}\right)^{-1}(0)\qquad(0\leqslant i,j\leqslant n+1)$$

is a smooth, compact 1-manifold; therefore, the classification theorem for 1-manifolds, see [Guillemin and Pollack, 1974], implies that each component (only a finite number) of $F(\{i,j\})$ is a smooth copy of the 1-sphere S^1; (2) the 1-manifolds $F(\{i,j\})$ intersect the smooth, compact n-manifolds $F_i^{-1}(0)$ or $F_j^{-1}(0)$ transversally, i.e.,

$$T_p(F(\{i,j\}))\oplus T_p(F_i^{-1}(0))=T_p\mathbb{R}^{n+1},\quad\text{for } p\in F(\{i,j\})\cap F_i^{-1}(0)$$

or

$$T_q(F(\{i,j\})) \oplus T_q(F_j^{-1}(0)) = T_q\mathbb{R}^{n+1}, \quad \text{for } q \in F(\{i,j\}) \cap F_j^{-1}(0),$$

respectively. Thus, the sets $F(\{i,j\}) \cap F_i^{-1}(0)$ and $F(\{i,j\}) \cap F_j^{-1}(0)$ consist of a finite number of points whose "algebraic" sum will be independent of the indices i and j.

In order to count these points in a suitable way the manifolds $F(\{i,j\})$ receive the following orientations: Recall that, in the case where $f: M \to N$ is smooth and $a \in N$ is a regular value for $f(f^{-1}(a) \neq \varnothing)$, the manifold $f^{-1}(a)$ inherits a natural orientation, provided M and N have an orientation, and this orientation is called the preimage orientation of $f^{-1}(a)$; see [Guillemin and Pollack, 1974].

Definition. Let $F = (F_0, \ldots, F_{n+1})$ be a regular function system and $0 \leqslant i, j \leqslant n+1, i \neq j$. Let (i,j) denote the following orientation of $F(\{i,j\})$:

$i<j$:

$n+1$	$i+j$	(i,j)
odd	odd	$-\omega$
odd	even	ω
even	odd	ω
even	even	$-\omega$

$$\omega := \text{preimage orientation of } \left(F_0, \ldots, \hat{F}_i, \ldots, \hat{F}_j, \ldots, F_{n+1}\right)^{-1}(0),$$

$$i>j\colon (i,j) := -(j,i)$$

The corresponding oriented manifolds $(F(\{i,j\}), (i,j))$ and $(F(\{i,j\}), (j,i))$ are denoted by $F(i,j)$, $F(j,i)$, respectively, in the following.

Now one can classify points from $F(\{i,j\}) \cap F_i^{-1}(0)$ or $F(\{i,j\}) \cap F_j^{-1}(0)$, respectively, in the following way; see [Kronecker, 1869a, p. 179]:

Definition. Let $F = (F_0, \ldots, F_{n+1})$ be a regular function system and $0 \leqslant i, j \leqslant n+1, i \neq j$. A point $e \in F(i,j) \cap F_i^{-1}(0)$ is called an **entrance point of $F(i,j)$** (into $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) < 0\}$), provided the following condition is satisfied:

Following the manifold $F(i,j)$ in the direction (i,j), one leaves the set $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) > 0\}$ at the point e and enters the set $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) < 0\}$.

$\boldsymbol{E(i,j)}$**:=** set of entrance points of $F(i,j)$.

A point $a \in F(i,j) \cap F_i^{-1}(0)$ is called an **exit point** of $F(i,j)$ (out of $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) < 0\}$), provided the following condition is satisfied:

Leaving the manifold $F(i,j)$ in the direction (i,j), one leaves the set $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) < 0\}$ at the point a and enters the set $\{x \in \mathbb{R}^{n+1}\colon F_j(x) \cdot F_i(x) > 0\}$.

$\boldsymbol{A(i,j)}$**:=** set of exit points of $F(i,j)$. (See Fig. 1.)

Now the definition of the Kronecker characteristic runs as follows [Kronecker, 1869a, p. 171]:

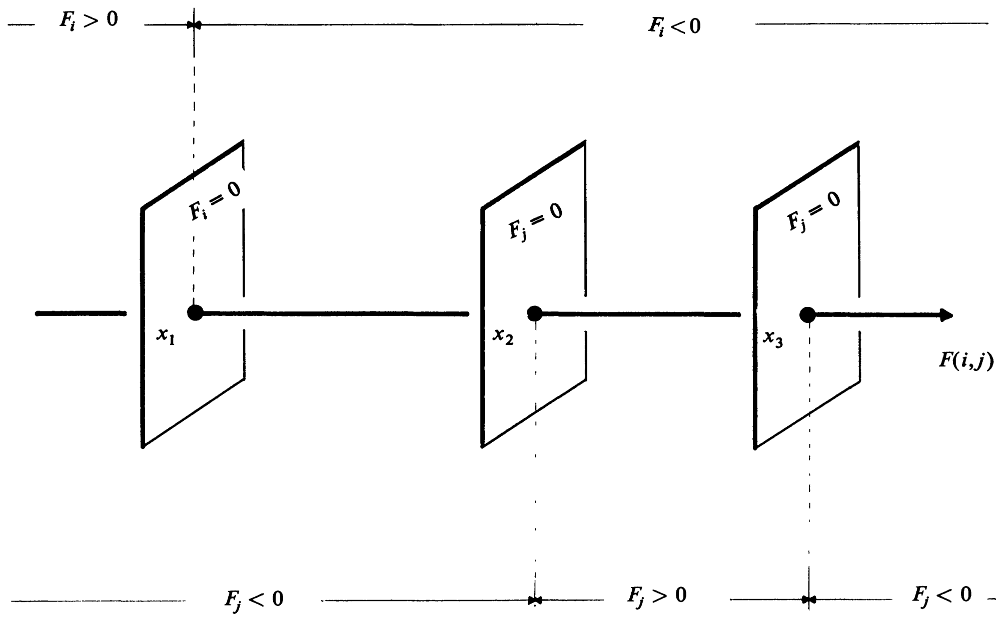
Theorem and Definition. *Let $F = (F_0, \ldots, F_{n+1})$ be a regular function system. Then the number*

$$\#E(i,j) - \#A(i,j)$$

is even and independent of the choice of the indices i and j.

The integer

$$\chi(F_0, \ldots, F_{n+1}) := 1/2(\#E(i,j) - \#A(i,j))$$



x_1 = exit point of $F(i,j)$ x_2 = exit point of $F(j,i)$ x_3 = entrance point of $F(j,i)$

FIG. 1

is called the Kronecker characteristic of the regular function system $F=(F_0,\dots,F_{n+1})$.

When proving this theorem [Kronecker, 1869a, p. 179–182] (see [Siegberg, 1977] for a modern proof), we can observe the following relationship:

$$\chi(F_0,\dots,F_{n+1}) = \sum_{\substack{(F_0,\dots,\hat{F}_j,\dots,F_{n+1})(x)=0\\ F_j(x)<0}} (-1)^j \operatorname{sign} d(F_0,\dots,\hat{F}_j,\dots,F_{n+1})(x)$$
$$= -\sum_{\substack{(F_0,\dots,\hat{F}_j,\dots,F_{n+1})(x)=0\\ F_j(x)>0}} (-1)^j \operatorname{sign} d(F_0,\dots,\hat{F}_j,\dots,F_{n+1})(x)$$
$$\text{for } 0 \leqslant j \leqslant n+1,$$

where the sign of a linear mapping is defined as follows.

Let E and F be finite dimensional real vector spaces with orientation, and let $f\colon E \to F$ be a linear isomorphism; then

$$\operatorname{sign} f := \begin{cases} +1, & f \text{ preserves orientation} \\ -1, & f \text{ reverses orientation.} \end{cases}$$

The identity above may be interpreted as the original Kronecker existence theorem:

KRONECKER EXISTENCE THEOREM. *Let* $F = (F_0,\dots,F_{n+1})$ *be a regular function system. Then* $\chi(F_0,\dots,F_{n+1}) \neq 0$ *implies that for each* j, $0 \leqslant j \leqslant n+1$, *the map* $(F_0,\dots,\hat{F}_j,\dots,F_{n+1})\colon \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ *has a zero in* $\{x \in \mathbb{R}^{n+1}\colon F_j(x) < 0\}$ *and also in* $\{x \in \mathbb{R}^{n+1}\colon F_j(x) > 0\}$.

Looking in the literature for notions related to the Kronecker characteristic or for theorems related to the Kronecker existence theorem, one discovers several papers of Gauss, which will be discussed in the following section.

3. The Fundamental Theorem of Algebra. Taking a closer look at Gauss's first [Gauss, 1799] and fourth [Gauss, 1850] proofs of the fundamental theorem of algebra, it turns out that these proofs contain essentially a more general version of the Kronecker existence theorem (in $\mathbb{R}^2$) (cf. also [Ostrowski, 1933 and 1934]).

Therefore we reformulate the Kronecker existence theorem in $\mathbb{R}^2$ in the following way:

THEOREM. *Let $F = (F_0, F_1, F_2) : \mathbb{R}^2 \to \mathbb{R}^3$ be a regular function system such that $F_0^{-1}(0)$ is connected. Fix an orientation of $F_0^{-1}(0)$, and let the intersection points $\{p_1, \ldots, p_s, p_{s+1} = p_1\} =: F(\{1,2\}) \cap F_1^{-1}(0)$ be indexed according to that fixed orientation.*

Label the points $p_1, \ldots, p_s$ as follows

$$l(p_j) := \begin{cases} 1(-1), & \text{if } F_2(p_j) > 0 \text{ and } j \text{ even (odd)} \\ -1(1), & \text{if } F_2(p_j) < 0 \text{ and } j \text{ even (odd)} \end{cases}$$

$$(1 \leqslant j \leqslant s).$$

Then $\Sigma_{j=1}^{s} l(p_j) \neq 0$ implies that the mapping

$$(F_1, F_2) : \mathbb{R}^2 \to \mathbb{R}^2$$

has a zero in $\{x \in \mathbb{R}^2 : F_0(x) < 0\}$ and also in $\{x \in \mathbb{R}^2 : F_0(x) > 0\}$.

Proof. We claim $|\chi(F_0, F_1, F_2)| = |\#E(1,2) - \#A(1,2)| = |\Sigma_{j=1}^{s} l(p_j)|$.

Now let us sketch Gauss's train of thought in his fourth proof [Gauss, 1850] of the fundamental theorem of algebra.

Let

$$P : \mathbb{C} \to \mathbb{C}, P(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n \qquad (a_n \neq 0),$$

be a polynomial with complex coefficients. Splitting P into its real and imaginary parts, $P(z) = U(z) + iT(z)$, we obtain smooth functions U and T such that the sets $U^{-1}(0)$, $T^{-1}(0)$ intersect a sphere S (centered at the origin) transversally in $2n$ points, provided the radius of S is sufficiently large. (See Fig. 2.)

At each intersection point $p \in U^{-1}(0) \cap S$ the map $U_{|S}$ changes sign, whereas the sign of $T_{|S}$ is constant ($\neq 0$) in a sufficiently small neighborhood of p. Labeling the intersection points $\{p_1, \ldots, p_{2n}\} =: U^{-1}(0) \cap S$ as in the previous theorem ($F_1 = U, F_2 = T$) it is easy to see that $|\Sigma_{j=1}^{2n} l(p_j)| = 2n$.

Now the existence of a root of P follows from the following argument.

Each component of $\{x \in \mathbb{R}^2 : U(x) < 0\}$ (or of $\{x \in \mathbb{R}^2 : U(x) > 0\}$) intersects the sphere S in one, two, or more open segments $(p_j p_{j+1})$ according to its shape; see Fig. 3. Therefore each point $p_j \in U^{-1}(0) \cap S$ whose index j is odd must be connected by a continuum in $U^{-1}(0)$ with a point $p_{\hat{j}} \in U^{-1}(0) \cap S$ whose index $\hat{j}$ is even [Gauss, 1850, p. 81–82]:

> Hence we have shown that among the n curves on which $U=0$ that start from a point of S with odd index in the interior [of S], some definite one traverses this interior continuously until it encounters a point with even index. [Gauss, 1913, p. 76–77]

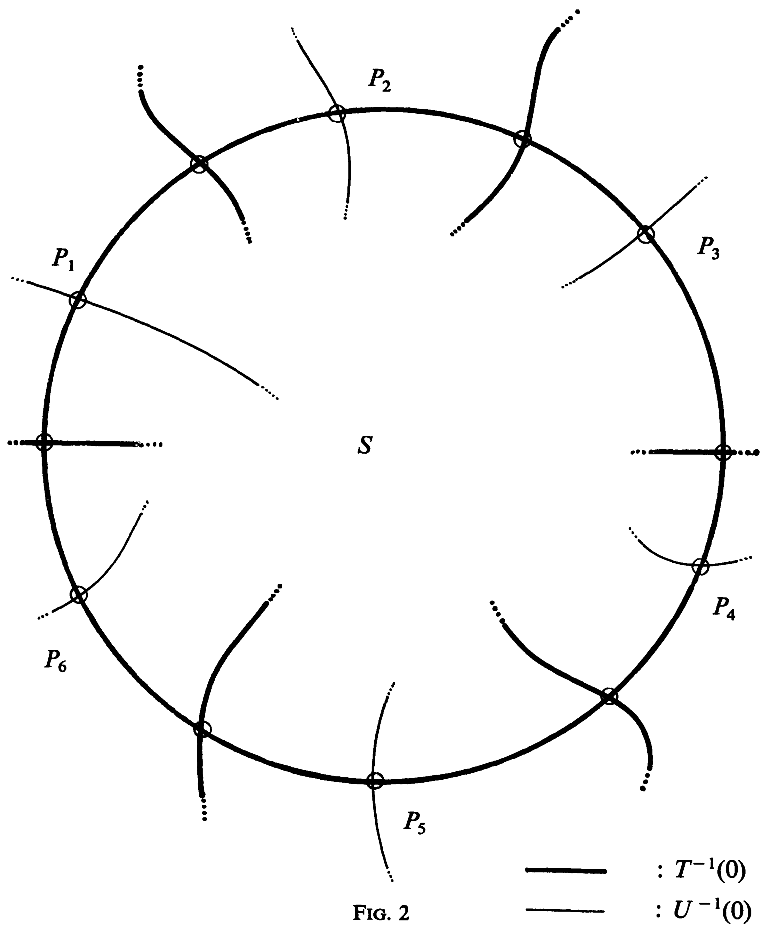
Because the sign of T at a point p_j with an odd index is different from the sign of T at a point $p_{\hat{j}}$ with an even index there must exist n zeros of P (not necessarily all distinct).

REMARKS. (i) From a modern viewpoint this argument of Gauss is not quite complete. The gap was filled in a note of Ostrowski [Ostrowski, 1933].

(ii) Gauss's argument described above can be improved in the following way [Ostrowski, 1934]:

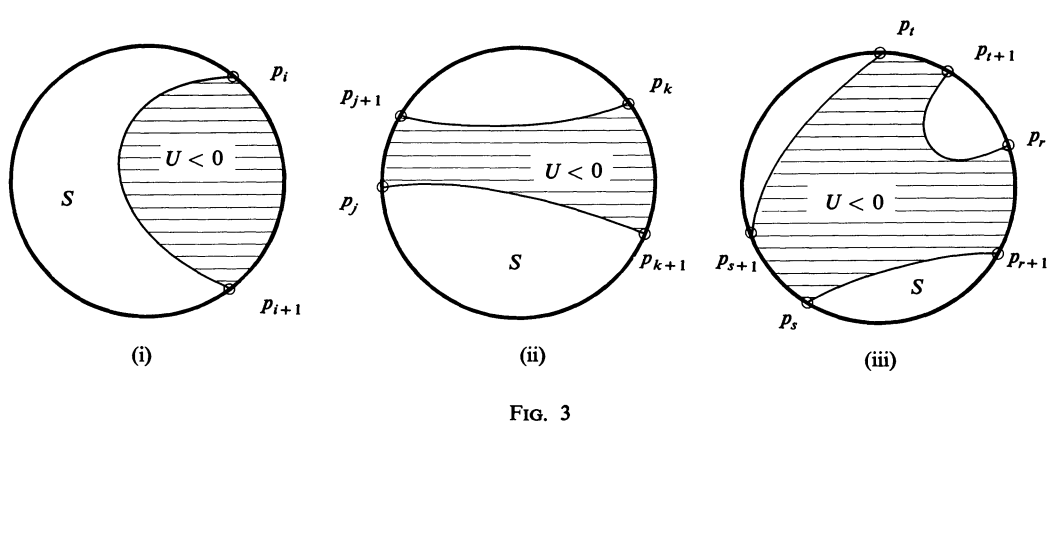
THEOREM. *Let $B := \{x \in \mathbb{R}^2 : \|x\| \leqslant R\}$, $S := \partial B$, be a ball in $\mathbb{R}^2$, and let $F = (F_1, F_2)$,*

$$F : (B, S) \to (\mathbb{R}^2, \mathbb{R}^2 - \{0\}),$$



P_1
P_2
P_3
P_4
P_5
P_6
S
: $T^{-1}(0)$
: $U^{-1}(0)$

FIG. 2



p_i
p_{i+1}
$U<0$
S
(i)
p_{j+1}
p_j
p_k
p_{k+1}
$U<0$
S
(ii)
p_t
p_{t+1}
p_r
p_{r+1}
p_{s+1}
p_s
$U<0$
S
(iii)

FIG. 3

be a continuous mapping with the following property:

> F_1 *has only a finite number of zeros on* S, *say* $p_1, \dots, p_{2n}, p_{2n+1} = p_1$ (*which are assumed to be indexed according to a fixed orientation of* S), *and* $F_{1|S}$ *changes its sign at each point* p_j, $1 \leqslant j \leqslant 2n$.

If $\Sigma_{j=1}^{2n} l(p_j) \neq 0$, *then* F *has a zero in* B.

Proof. (cf. [Ostrowski, 1934]). Each p_j is connected by a continuum in $F_1^{-1}(0)$ with a point $p_{\hat{j}}$ such that $j \not\equiv \hat{j} \bmod 2$ (proof!). Since $\Sigma_{j=1}^{2n} l(p_j) \neq 0$, one concludes that F_2 has at least one zero on at least one component of $F_1^{-1}(0)$.

Gauss's fourth proof [Gauss, 1850] of the fundamental theorem of algebra is only a slightly different version of his first proof (and dissertation) [Gauss, 1799]. There Gauss considers only the sets $U^{-1}(0)$, $T^{-1}(0)$ rather than the components of $\{x \in \mathbb{R}^2 : U(x) \lesseqgtr 0\}$ as described above, and claims that there must be an intersection point of $U^{-1}(0)$ and $T^{-1}(0)$ (assuming that P is a polynomial with real coefficients). Also this proof contains a gap, because Gauss uses without proof a classification for algebraic curves which states—roughly speaking—that each branch of an algebraic curve is isomorphic either to a circle or to an interval [Gauss, 1799, p. 27]:

> It is known from higher geometry that every algebraic curve [here $U^{-1}(0)$, $T^{-1}(0)$] (or every part of an algebraic curve, if the curve happens to have several parts) either repeats itself or extends to infinity in both directions, and therefore if a branch of an algebraic curve enters a bounded region it must necessarily leave again. [Gauss, 1913, p. 33]

A precise proof can be found, again, in [Ostrowski, 1933].

In view of the Kronecker integral, which in the 1-dimensional case is exactly the representation of the winding number in complex function theory, Gauss's third proof [Gauss, 1816] should be mentioned. Taking a closer look at the relation $|\Sigma_{j=1}^{2n} l(p_j)| = 2n$, one observes that it describes the fact that the polynomial $P_{|S}$ winds n times around the origin. ($\Sigma_{j=1}^{2n} l(p_j)$ may be interpreted as the intersection number of $P(S)$ with $0 \times \mathbb{R}$.)

From his correspondence with Bessel, one knows that Gauss was acquainted with the foundations of complex function theory and that he knew about the meaning of complex logarithm. In a letter to Bessel (dated from Dec. 18, 1811) [Gauss, 1811], Gauss reports about complex integration and the meaning of the line integral $\int_\gamma \phi(x)dx$. He explains how a complex logarithm should be defined ($\log(x) = \int_1^x (1/x)dx$), and, furthermore, he observes that the logarithm changes its value according to how the line from 1 to x rotates around the origin:

> It is clear how a function generated by $\int \phi(x)dx$ can have many values for a single value of x, since the path of integration can go around a point where $\phi(x) = \infty$ either not at all, or once, or several times. For example, if we define $\log x$ by $\int (1/x)dx$, starting from $x = 1$, we obtain $\log(x)$ either without going around the point $x = 0$, or after going around it once or several times; each time there is an additional $+2\pi i$ or $-2\pi i$: this makes the many logarithms of each number perfectly obvious. [Gauss, 1811]

Thus, on account of these lines one might suspect that Gauss interpreted the relation $|\Sigma_{j=1}^{2n} l(p_j)| = 2n$ in a "correct" (function theoretic) way and that he tried to prove the fundamental theorem via complex function theory. Indeed, in a lecture, "Theorie der imaginären Grössen (1840)," Gauss mentioned (see [Fraenkel, 1922]) that his third proof of the fundamental theorem of algebra [Gauss, 1816] originated from his first one, and he gave the function-theoretic argument that the winding number $w(P_{|S}, 0)$ equals n, whereas the winding number of any map $F:(B, S) \to (\mathbb{R}^2, \mathbb{R}^2 - \{0\})$ vanishes if there is no zero of F in B; see [Fraenkel, 1922].

However, this argument cannot be found explicitly in [Gauss, 1816]. It is shown there, with careful avoidance of any complex function theory, that the expression

$$\int_0^{2\pi} \int_0^R \frac{\partial}{\partial r} \left[\frac{U \dfrac{\partial T}{\partial \phi} - T \dfrac{\partial U}{\partial \phi}}{U^2 + T^2} \right] dr d\phi$$

($z = re^{i\phi}$; R sufficiently large) has two different values, according to the order of integration:

$$\int_0^{2\pi}\left[\int_0^R \frac{\partial}{\partial r}\left\{\frac{U\dfrac{\partial T}{\partial \phi} - T\dfrac{\partial U}{\partial \phi}}{U^2 + T^2}\right\} dr\right] d\phi > 0 \tag{*}$$

and

$$\int_0^{2\pi}\left[\int_0^R \frac{\partial}{\partial r}\left\{\frac{U\dfrac{\partial T}{\partial \phi} - T\dfrac{\partial U}{\partial \phi}}{U^2 + T^2}\right\} d\phi\right] dr = 0.$$

On the other hand, Gauss proves that both integrals must coincide (assuming that the polynomial $P = U + iT$ has no roots).

$$\left\{\text{Observe:} \frac{\partial}{\partial r}\left\{\frac{U\dfrac{\partial T}{\partial \phi} - T\dfrac{\partial U}{\partial \phi}}{U^2 + T^2}\right\} = \frac{\partial^2 \arg P(z)}{\partial r\, \partial \phi}.\right\}$$

This contradiction proves the theorem.

Complex function theory shows through only in a footnote (see [Gauss, 1816, p. 62]; [Gauss, 1913, p. 65]), when it is mentioned that one can show in a different way that the integral (*) equals $2n\pi$.

Summarizing these observations we are tempted to say, in accordance with [Fraenkel, 1922], that Gauss's first and third proofs of the fundamental theorem of algebra were based on the same principle, which today is called the argument principle (or Kronecker's existence theorem).

REMARK. As Fraenkel reports in [Fraenkel, 1922], the argument principle was first used (explicitly) for a proof of the fundamental theorem of algebra in Briot and Bouquet's "Théorie des fonctions elliptiques" (2nd ed., I, Paris, 1873, p. 20). The argument principle seems to be first stated (and proved) by Cauchy in his so-called "Turin Transactions" (1831), see [Osgood, 1907, p. 184]; see also [Cauchy, 1835].

4. The Kronecker Integral. After he had completed his theory of characteristics (as described in section 2) Kronecker's attention was directed by Weierstrass to the complex function-theoretic or potential-theoretic aspect of the characteristic [Kronecker, 1869a, p. 205]. In particular, the representation of the Kronecker characteristic by an integral seems to have been influenced by discussions with Weierstrass [Kronecker 1869a, p. 205–206]. The following relation is the key of the connection between the Kronecker characteristic and the Kronecker integral [Kronecker, 1869a, p. 186].

THEOREM. *Let $F = (F_0, \dots, F_{n+1})$ be a regular function system and $0 \leqslant i, j \leqslant n + 1$, $i \neq j$. Let Φ_{ij} be the mapping*

$$\Phi_{ij}\colon F(i,j) \to \mathbb{R}^2 - \{0\}$$
$$x \mapsto (F_i(x), F_j(x)).$$

Then

$$\chi(F_0, \dots, F_{n+1}) = w(\Phi_{ij}, 0) \qquad (\textit{winding number of } \Phi_{ij} \textit{ around the origin}).$$

Proof. The proof uses the representation of the winding number $w(\Phi_{ij}, 0)$ via the intersection number of ϕ_{ij} with $0 \times \mathbb{R}$.

Roughly speaking the theorem means the following: If $i,j \neq 0$ and $i \neq j$, then the (x_i, x_j)-plane in $\mathbb{R}^{n+1}$ intersects the set $(F_1, \ldots, F_{n+1})(F_0^{-1}(0))$ precisely at $\Phi_{ij}(F(i,j))$ (assume $\Phi_{ij}(F(i,j))$ to be a subset of the (x_i, x_j)-plane). Thus the theorem means that the image $(F_1, \ldots, F_{n+1})(F_0^{-1}(0))$ winds exactly $\chi(F_0, \ldots, F_{n+1})$ times around the origin in the (x_i, x_j)-plane for each $i,j \neq 0$ and $i \neq j$.

Because of this observation Kronecker was led "immediately" (as he writes, see [Kronecker, 1869, p. 187]) to a representation of the characteristic by an integral:

$$\chi(F_0, \ldots, F_{n+1}) = (-1)^j (\operatorname{vol} S^n)^{-1} \int_{F_j^{-1}(0)} \left(r \circ \left(F_0, \ldots, \hat{F}_j, \ldots, F_{n+1}\right)\right)^* \sigma$$

for each $0 \leqslant j \leqslant n+1$. (In addition, it is assumed that $\{x \in \mathbb{R}^{n+1} : F_j(x) \leqslant 0\}$ or $\{x \in \mathbb{R}^{n+1} : F_j(x) \geqslant 0\}$ is compact and that $F_j^{-1}(0)$ is oriented by the preimage orientation.)

Now the Kronecker existence theorem has the following more familiar form:

Kronecker Existence Theorem. *Let $F = (F_0, \ldots, F_{n+1})$ be a regular function system. Then*

$$(\operatorname{vol} S^n)^{-1} \int_{F_j^{-1}(0)} \left(r \circ \left(F_0, \ldots, \hat{F}_j, \ldots, F_{n+1}\right)\right)^* \sigma \neq 0$$

implies that the map $(F_0, \ldots, \hat{F}_j, \ldots, F_{n+1}) : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ has a zero in $\{x \in \mathbb{R}^{n+1} : F_j(x) < 0\}$ and also in $\{x \in \mathbb{R}^{n+1} : F_j(x) > 0\}$.

Remark. Using arguments of potential theory or Stokes's theorem, the theorem can be improved in the following way [Kronecker, 1869, ch. VI and ch. VIII]:

Let $F_0 : \mathbb{R}^{n+1} \to \mathbb{R}$ be a smooth map such that $0 \in \mathbb{R}$ is a regular value for F_0 and $F_0^{-1}((-\infty, 0])$ is compact. Let

$$(F_1, \ldots, F_{n+1}) : F_0^{-1}((-\infty, 0]) \to \mathbb{R}^{n+1}$$

be a smooth mapping such that $0 \notin (F_1, \ldots, F_{n+1})(F_0^{-1}(0))$. Then $K((F_1, \ldots, F_{n+1})|_{F_0^{-1}(0)}) \neq 0$ implies that the mapping $(F_1, \ldots, F_{n+1})$ has a zero.

(Clearly, $F_0^{-1}((-\infty, 0])$ can be replaced by any compact oriented smooth manifold with boundary.)

As already mentioned in section 3, the Kronecker integral coincides with the well-known integral representation of the winding number in complex function theory. But also in the higher dimensional case "embryonic" forerunners of the Kronecker integral exist. Here we have to mention Gauss again. A remark of Kronecker [Kronecker, 1869, p. 190] shows that a special case of the Kronecker integral can be found in the early work of Gauss in potential theory [Gauss, 1813, p. 9]:

Let $v(x) := (x - x_0)/\|x - x_0\|^3 \in \mathbb{R}^3$ be the electric field due to a point charge 1 at x_0. Then the **flux of v across a closed surface** $S \subset \mathbb{R}^3$ is defined by

$$\text{flux across } S := \int_S \langle v, N \rangle \, ds,$$

where N is the outward normal and ds the area element. A theorem of Gauss [Gauss, 1813, p. 9] says that the flux of v across S equals 4π or vanishes according as the point x_0 lies inside of S or not. Translating this, one observes that the integral describing the flux corresponds exactly to the Kronecker integral of the inclusion $i : S \to \mathbb{R}^3 - \{x_0\}$ (times 4π):

$$\int_S \langle v, N \rangle \, ds = \int_S (r \circ (i - x_0))^* \sigma.$$

There is another prominent forerunner of the Kronecker integral—the **Gaussian linking** number [Gauss, 1833]:

Let $f, g:[0,1]\to\mathbb{R}^3$ be two smooth loops such that $f(s)\neq g(t)$ for all $s, t\in[0,1]$. Then the **linking number of f and g** is defined by

$$\operatorname{link}(f,g) = -(4\pi)^{-1}\int_0^1\int_0^1 \|f(s)-g(t)\|^{-3}\det\big(f(s)-g(t),\dot f(s),\dot g(t)\big)\,ds\,dt.$$

It is easy to see that link(f,g) is exactly the winding number of the mapping $F: S^1\times S^1\to\mathbb{R}^3-\{0\}$, $F(x,y)=\tilde f(x)-\tilde g(y)$ ($\tilde f$ and $\tilde g$ denote the maps corresponding to f and g via $\exp(2\pi i\lambda)$), and that the integral describing the linking number corresponds exactly to the Kronecker integral $K(F)$ of F. Again the linking number has a physical (electrodynamical) meaning: If f and g parametrize closed loops M and N, respectively, then the magnetic field at $x\in M$, due to a steady unit electric current flowing around the loop N is given by

$$G(x) = -(4\pi)^{-1}\int_N \frac{(y-x)\times dy}{\|x-y\|^3}.$$

It follows that link$(f,g)=\int_M G(x)\,dx$ is precisely the work done by this field on a unit magnetic pole which makes one circuit of M, see [Flanders, 1962].

REMARK. It is interesting to see that Bohl, who proved in 1904 [Bohl, 1904] that there exists no (C^1-) retraction of the cube onto its boundary (he proved also the Kronecker existence theorem), using the Kronecker integral and Stokes's theorem, apparently did not know the work of Kronecker and subsequent papers of Poincaré [Poincaré, 1886] and Picard [Picard, 1892] about the Kronecker integral: he refers to a "certain" integral, see [Bohl, 1904, ch. 2].

5. Sturm Chains and the Kronecker Characteristic. It seems that the motivation for Kronecker when he began his paper [Kronecker, 1869a] was to generalize the theorem of Sturm:

> In the investigations whose development I have outlined here, I started from Sturm's theorem. An extension of this theorem to systems of equations was given long ago by Hermite [Comptes rendus t. 35, 1852, II, p. 52; Comptes rendus t. 36, 1853, I, p. 294]; but I was able to generalize the continued fraction algorithm on which Sturm's work was based, and then I could see a natural interpretation of the results so obtained. [Kronecker, 1869a, p. 205]

In his papers [Sturm, 1829 and 1835] Sturm succeeded for the first time in finding the exact number of distinct real roots of a polynomial with real coefficients. Since that time there have been many efforts to generalize Sturm's theorem.

The following generalized version of **Sturm's theorem**, due to Sylvester [Sylvester, 1853], seems to have played an important role in Kronecker's considerations; see [Kronecker, 1869a, ch. III].

THEOREM. *Let p_1 and p_2 be two real polynomials (not identically zero) such that* $\deg p_1>\deg p_2$. *Let $a<b$ be real numbers such that $p_1(a)p_1(b)\neq 0$ and such that the roots of p_1 in (a,b) are simple. Further, let $p_3,\dots,p_n$ be polynomials such that*

$$p_1=q_2p_2-p_3,\ p_2=q_3p_3-p_4,\dots,p_{n-1}=q_np_n$$

for certain polynomials $q_2,\dots,q_n$ ("Euclidean algorithm").

If $p_n(x)\neq 0$ for $x\in[a,b]$, then the following identity holds:

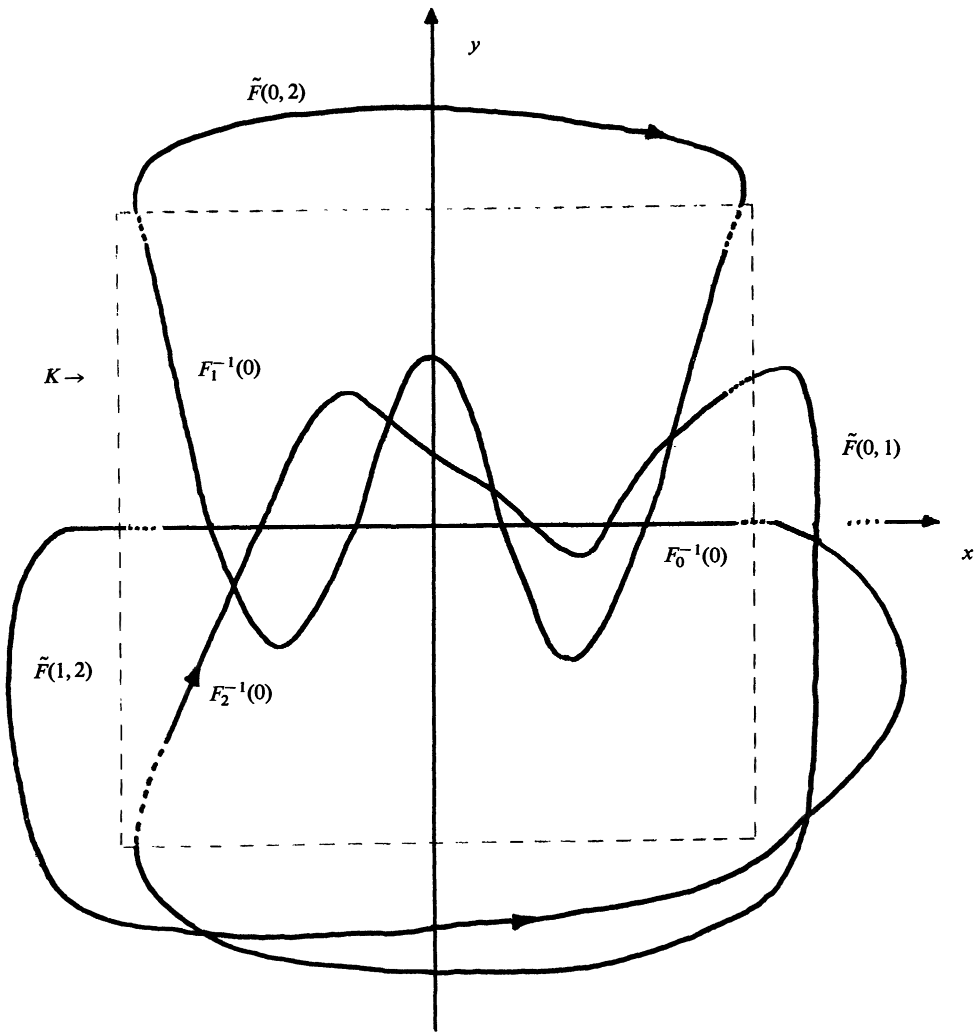
$$\#\{x\in(a,b): p_1(x)=0 \text{ and } p_1'(x)p_2(x)>0\}-\#\{x\in(a,b): p_1(x)=0 \text{ and } p_1'(x)p_2(x)<0\}$$
$$=V(p_1(a),\dots,p_n(a))-V(p_1(b),\dots,p_n(b)).$$

(V means the number of variations of sign in the chain $p_1,\dots,p_n$.)

This theorem can be embedded into the theory of characteristics as follows, see [Kronecker, 1869a, ch. IV] and [Kronecker, 1873, ch. I and ch. VIII]:

Let $F_i: \mathbb{R}^2 \to \mathbb{R}$ $(i=0,1,2)$ be the maps $F_1(x,y)=p_1(x)-y$, $F_2(x,y)=p_2(x)-y$ and $F_0(x,y)=y$.

Think of $F=(F_0,F_1,F_2)$ as a regular function system $\tilde{F}$; i.e., find a cube $K \subset \mathbb{R}^2$ which contains all intersection points $F_1^{-1}(0) \cap F_2^{-1}(0)$, $F_0^{-1}(0) \cap F_2^{-1}(0)$ and $F_0^{-1}(0) \cap F_1^{-1}(0)$, and replace the pieces $F_i^{-1}(0) \setminus K (i=0,1,2)$ by curves such that $\tilde{F}_i^{-1}(0)$ are loops for $i=0,1,2$, and such that $F_{|K}=\tilde{F}_{|K}$. (See Fig. 4.)



The function system $\tilde{F}=(\tilde{F}_0,\tilde{F}_1,\tilde{F}_2)$

FIG. 4

This can be done in such a way that, in the case where $\deg p_1$(or $\deg p_2$) is even, no additional intersection points of the "new" loop $\tilde{F}_1^{-1}(0)$ (or $\tilde{F}_2^{-1}(0)$) with the x-axis occur; in the case where $\deg p_1$ (or $\deg p_2$) is odd, this can be done in such a way that only one additional intersection point with the x-axis occurs. Now the following identity holds:

$$V(p_1(-\infty),\ldots,p_n(-\infty)) - V(p_1(\infty),\ldots,p_n(\infty))$$
$$= \begin{cases} \#E(1,2) - \#A(1,2) = 2\chi(F_0,F_1,F_2), & \deg p_1 \text{ even} \\ \#E(1,2) - \#A(1,2) = 2\chi(F_0,F_1,F_2) \pm 1, & \deg p_1 \text{ odd.} \end{cases}$$

As Kronecker writes [Kronecker, 1869a, p. 185], [Kronecker, 1863, ch. I, p. 306], this relation had been investigated by Sylvester, who conjectured a higher dimensional generalization of this fact.

It is curious to see how the "algebraist" Kronecker comments on his generalization of Sylvester's ideas. He remarks that the difficulties of generalization (for Sylvester) seem to be caused by Sylvester's using only algebraic tools and no geometric considerations.

> The simple case mentioned here was, as far as I know, first discussed by Sylvester, who interpreted it in a similar but less intuitive way.... The difficulties about the generalization that Sylvester mentions appear to come mostly from his restricting his attention to algebraic structures. As soon as I saw that all the relevant considerations belong exclusively to the more general domains which are considered in the "Geometry of position" when $n = 2$ or 3, I found very simple means of overcoming the difficulties that confronted me. [Kronecker, 1869a, p. 185]

6. Curvatura Integra. The **global Gauss-Bonnet theorem** for a compact orientable 2-dimensional Riemannian manifold M states that

$$\int_M K\,dv = 2\pi\chi(M),$$

where K is the **Gaussian curvature** of the surface M, dv is the **area element** of M, and $\chi(M)$ is the **Euler number** of M. From the Gaussian definition of curvature, it follows that in the case where M is a hypersurface in $\mathbb{R}^3$ the differential form $K\,dv$ equals $g^*\sigma$, where $g\colon M \to S^2$ is the Gauss map and σ is the volume form on S^2.

Therefore, in this case the Gauss-Bonnet theorem can be formulated equivalently [Guillemin and Pollack, 1974]: The winding number $w(g,0)(=\deg(g))$ of the Gauss map g equals $(1/2)\chi(M)$. The winding number of the Gauss map is often called the "**curvatura integra**" (cf. [Hopf, 1925]).

Part II of Kronecker's investigations on the characteristic of function systems [Kronecker, 1869b] deals with the generalization of the Gauss-Bonnet theorem. The object of his investigations is the characteristic of the function system $(F_0, F_{01},\ldots, F_{0n})$, where $F_0\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth map such that $F_0^{-1}((-\infty,0])$ is compact, $0 \in \mathbb{R}$ is a regular value for F_0 and where $F_{0i} = \partial F_0/\partial x_i$. Here, by the characteristic of $(F_0, F_{01},\ldots, F_{0n})$ Kronecker means the Kronecker integral:

$$\chi(F_0, F_{01},\ldots, F_{0n}) = (\operatorname{vol} S^{n-1})^{-1}\int_{M:=\partial F_0^{-1}((-\infty,0])} (r\circ(F_{01},\ldots, F_{0n}))^*\sigma$$

The map $r\circ(F_{01},\ldots, F_{0n})$ can be interpreted as the $(n-1)$-dimensional Gauss map $g\colon M \to S^{n-1}$, and again there is a smooth function $K\colon M \to \mathbb{R}$ such that $K\,dv = (r\circ(F_{01},\ldots, F_{0n}))^*\sigma$.

In [Kronecker, 1869b] Kronecker shows that the 2-dimensional theory of curvature, which was introduced by Gauss, has a natural generalization to higher dimensions (for hypersurfaces). In particular, the function K has a representation $K = k_1,\ldots, k_{n-1}$, where the k_i are generalized principal curvatures (cf. [Kobayashi and Nomizu, 1963, p. 33]). Hence, in the case where $0 \in \mathbb{R}^n$ is a regular value for $(F_{01},\ldots, F_{0n})$ one can compute the curvatura integra of $M = \partial F_0^{-1}((-\infty,0])$ as follows:

$$(\operatorname{vol} S^{n-1})^{-1}\int_M K\,dv = (\operatorname{vol} S^{n-1})^{-1}\int_M (r\circ(F_{01},\ldots, F_{0n}))^*\sigma$$
$$= \sum_{\substack{(F_{01},\ldots, F_{0n})(x)=0 \\ F_0(x)<0}} \operatorname{sign} d(F_{01},\ldots, F_{0n})(x).$$

An actual computation (implicitly) of the curvatura integra in the 2-dimensional case can be found in a paper of Poincaré [Poincaré, 1886]. Being interested in the location of equilibrium points for the ordinary differential equation

$$\frac{dx}{dt}=X, \quad \frac{dy}{dt}=Y, \quad \frac{dz}{dt}=Z \qquad (X, Y, Z \text{ polynomials of } x, y, z)$$

he looks for surfaces $F_0^{-1}(0)$ ($F_0: \mathbb{R}^3 \to \mathbb{R}, 0 \in \mathbb{R}$ regular value for F_0) **"without contact"** to the vector field, which means

$$\frac{\partial F_0}{\partial x}\cdot X+\frac{\partial F_0}{\partial y}\cdot Y+\frac{\partial F_0}{\partial z}\cdot Z<0\ (>0) \text{ on } F_0^{-1}(0). \tag{$*$}$$

He shows that the characteristic $\chi(F_0, X, Y, Z)$ (strictly speaking:

$$(\operatorname{vol} S^2)^{-1}\int_{F_0^{-1}(0)}(r\circ(X,Y,Z))^*\sigma)$$

equals the characteristic $\chi(F_0, F_{01}, F_{02}, F_{03})$ in the case where $(*)<0$, and equals $\chi(F_0, -F_{01}, -F_{02}, -F_{03})$ in the case where $(*)<0$.

Poincaré's proof is essentially a proof of the homotopy invariance of the winding number (**"Poincaré-Bohl" principle**). (However, it should be noted that Kronecker also gave a proof of the homotopy invariance of the characteristic, see [Kronecker, 1878].) Consequently, if $F_0^{-1}(0)$ is a surface without contact, the $\chi(F_0, F_{01}, F_{02}, F_{03})\neq 0$ implies the existence of an equilibrium point.

Poincaré computes $\chi(F_0, F_{01}, F_{02}, F_{03})$ for the case that $F_0^{-1}(0)$ is a surface of genus zero or one, and then gives the following general formula:

$$\chi(F_0, F_{01}, F_{02}, F_{03})=1-p, \qquad \text{if } F_0^{-1}(0) \text{ is a surface of genus } p$$
$$(=(1/2)\chi(F_0^{-1}(0))).$$

Thus, we have the Gauss-Bonnet formula.

Concerning the higher dimensional case there are only indications in Poincaré's paper: he conjectures that the characteristic $\chi(F_0, F_{01},\ldots, F_{0n})$ is a topological invariant:

> An $(n-1)$-dimensional manifold will be characterized, from the point of view of analysis situs, by its $n-2$ orders of connection as defined by Riemann and by Brioschi [Poincaré surely intends Betti]. The index of multiplicity of a manifold without contact $[=\chi(F_0, F_{01},\ldots, F_{0n})]$ will depend only on its orders of connection. [Poincaré, 1886, p. 186]

These suggestions have been taken up and made precise in the subsequent work of v. Dyck [v. Dyck, 1888 and 1890]. In modern notation v. Dyck's work can be described as follows.

Each compact smooth manifold M (with boundary) is homotopy equivalent to a finite CW-complex. If the number of q-cells ($0\leqslant q\leqslant n:=\dim M$) into which M is composed is denoted by α_q, then the following identity for the Euler number holds:

$$\chi(M)=\sum_{q=0}^{n}(-1)^q\alpha_q.$$

V. Dyck obtains this sum by means of gluing and cutting procedures (which remind one of Morse theory). If $M=F_0^{-1}((-\infty,0])$ and $0\in\mathbb{R}^n$ is a regular value for $(F_{01},\ldots,F_{0n})$, then $F_0: F_0^{-1}((-\infty,0])\to(-\infty,0]$ is a Morse function on $F_0^{-1}((-\infty,0])$, and Morse theory tells us that $F_0^{-1}((-\infty,0])$ is homotopy equivalent to a CW-complex consisting of exactly $\nu_q(F_0)$ ($0\leqslant q\leqslant n$) q-cells, where $\nu_q(F_0)$ is the number of (nondegenerate) **critical points of F_0 with Morse index q.** Therefore,

$$\chi(M)=\sum_{q=0}^{n}(-1)^q\nu_q(F_0)$$

$$= \sum_{\substack{(F_{01},\dots,F_{0n})(x)=0 \\ F_0(x)<0}} \operatorname{sign} d(F_{01},\dots,F_{0n})(x).$$

(Observe that sign $d(F_{01},\dots,F_{0n})(x)=(-1)^{\text{Morse index of } x}$.)

Finally, the Euler characteristic satisfies the relation

$$\chi(\partial M)=\chi(M)+(-1)^{\dim \partial M}\chi(M);$$

and, if n is odd, this implies the $(n-1)$-dimensional Gauss-Bonnet theorem (for hypersurfaces $F_0^{-1}(0)$):

$$\int_{F_0^{-1}(0)} K\,dv=(\operatorname{vol} S^{n-1})\chi(F_0,F_{01},\dots,F_{0n})$$

$$=(1/2)(\operatorname{vol} S^{n-1})\chi(F_0^{-1}(0)).$$

REMARK. A complete proof of the Gauss-Bonnet theorem for hypersurfaces $M\subset\mathbb{R}^n$ (n odd) was given by Hopf in his dissertation (Berlin, 1935) [Hopf, 1925].

I would like to thank Prof. Dr. Heinz-Otto Peitgen for his encouragement and his advice during the preparation of this paper.

References

P. Alexandroff and H. Hopf, 1935: Topologie, Springer, Berlin; corrected reprint, 1974.

P. Bohl, 1904: Über die Bewegung eines mechanischen Systems in der Nähe einer Gleichgewichtslage, J. Reine Angew. Math. 127, 179–276.

L. E. J. Brouwer, 1911: Beweis der Invarianz der Dimensionszahl, Math. Ann. 70, 161–165.

______, 1912: Über Abbildung von Mannigfaltigkeiten, Math. Ann. 71, 97–115.

A. L. Cauchy, 1835: Calcul des indices des fonctions, J. de l'école polyt. t. 15, xxv-ième, 176–229.

W. v. Dyck, 1888: Beiträge zur Analysis Situs I, Math. Ann. 32, 457–512.

______, 1890: Beiträge zur Analysis Situs II, Math. Ann. 37, 273–316.

H. Flanders, 1962: Differential Forms, Academic Press.

A. Fraenkel, 1922: Der Zusammenhang zwischen dem ersten und dem dritten Gausschen Beweis des Fundamentalsatzes der Algebra, Jahresber. der DMV 31, 234–238.

C. F. Gauss, 1799: Gesammelte Werke, Bd. III, Gesellschaft der Wissenschaften zu Göttingen, 1–31.

______, 1811: Gesammelte Werke, Bd. VIII, Gesellschaft der Wissenschaften zu Göttingen, 90–92, and Gesammelte Werke, Bd. 10, 1, 365–371.

______, 1813: Gesammelte Werke, Bd. V, Gesellschaft der Wissenschaften zu Göttingen, 1–22.

______, 1816: Gesammelte Werke, Bd. III, Gesellschaft der Wissenschaften zu Göttingen, 57–64.

______, 1833: Gesammelte Werke, Bd. V, Gesellschaft der Wissenschaften zu Göttingen, 605.

______, 1850: Gesammelte Werke, Bd. III, Gesellschaft der Wissenschaften zu Göttingen, 72–103.

______, 1913: Die vier Beweise für die Zerlegung ganzer algebraischer Funktionen, Ostw. Klass. Nr. 14, (3. Auflage).

V. Guillemin and A. Pollack, 1974: Differential Topology, Prentice-Hall, Englewood Cliffs, N.J.

J. Hadamard, 1910: Sur quelques applications de l'indice de Kronecker, in J. Tannery: Introduction à la théorie des fonctions d'une variable, 2ème ed., vol 2, Hermann, Paris.

H. Hopf, 1925: Über die Curvatura Integra geschlossener Hyperflächen, Math. Ann. 95, 340–367.

S. Kobayashi and K. Nomizu, 1963: Foundations of Differential Geometry, vol 2, New York: Interscience (Wiley), New York.

L. Kronecker, 1869a: Über Systeme von Functionen mehrer Variabeln, Monatsberichte der Königlich Preussischen Akademie der Wissenschaften, 159–193, in: Gesammelte Werke, Bd. I, Teubner, Leipzig, 177–212.

______, 1869b: Über Systeme von Functionen mehrer Variabeln, Monatsberichte der Königlich Preussischen Akademie der Wissenschaften, 688–698, in: Gesammelte Werke, Bd. I, Teubner, Leipzig, 213–226.

______, 1873: Über die verschiedenen Sturm'schen Reihen und ihre gegenseitigen Beziehungen, Monatsberichte der Königlich Preussischen Akademie der Wissenschaften, 117–154, in: Gesammelte Werke, Bd. I, Teubner, Leipzig, 304–348.

______, 1878: Über die Charakteristik von Functionen-Systemen, Monatsberichte der Königlich Preussischen Akademie der Wissenschaften, 145–152, in: Gesammelte Werke, Bd. II, Teubner, Leipzig, 71–82.

W. F. Osgood, 1907: Lehrbuch der Funktionentheorie, Teubner, Leipzig.

A. Ostrowski, 1933: Über den ersten und vierten Gausschen Beweis des Fundamentalsatzes der Algebra, Materialien zu einer wissenschaftlichen Biographie von Gauss, in: Gauss, Gesammelte Werke, Bd. 10, 2, Abh. 3.

______, 1934: Über Nullstellen stetiger Funktionen zweier Variabeln, J. Reine Angew. Math. 170, 83–94.

H. Poincaré, 1886: Sur les courbes définies par une équation différentielle IV, J. Math. Pures Appl. (4) 2, 151–217.

É. Picard, 1892: Sur le nombre des racines communes à plusieurs équations simultanées, J. Math. Pures Appl. (4) 8, 5–24.

H. W. Siegberg, 1977: Abbildungsgrade in Analysis und Topologie, Diplomarbeit, Universität Bonn.

C. F. Sturm, 1829: Analyse d'un Mémoire sur la résolution des équations numériques, Bull. des Sciences de Férussac XI, 419.

______, 1835: Mémoire sur la résolution des équations numériques, Mém. des savants étrangers VI, 271.

J. J. Sylvester, 1853: On a theory of the syzygetic relations of two rational integral functions, comprising an application to the theory of Sturm's functions, Philos. Trans. Roy. Soc. London, 143.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

CHARACTERIZATION OF CLASSES OF FUNCTIONS

R. CREIGHTON BUCK

Department of Mathematics, University of Wisconsin, Madison, WI 53706

Some areas of elementary analysis abound with interesting unsolved research problems. The purpose of this note is to examine several of these. All are related indirectly to the 13th Problem in Hilbert's famous list. As originally formulated, this dealt with formulas for the solution of a seventh degree polynomial but was quickly recast as the conjecture that there exist continuous functions of three variables that are not expressible in terms of continuous functions of two variables. The astonishing negative solutions of Kolmogorov and Arnold in 1957 showed that, on the contrary, every continuous function of n variables could be expressed as a superposition of continuous functions of *one* variable and the single binary function $+$. (See [1] and [3] for additional background.)

The questions to be discussed in this note deal with other superposition problems in which the representation scheme is restricted.

Equivalent descriptions of a mathematical object may often be operationally quite distinct. For example, a description of a class that may be quite suitable for distinguishing this class from others may not be at all helpful in testing a specific object for membership in that class. As an illustration consider the following:

DEFINITION 1. *Choose an integer $p>0$, and let $\mathfrak{F}^p$ be the smallest subclass of $C[R^3]$ with the property that, for any functions $F\in\mathfrak{F}^p$ and $g\in C^p[R]$, each of the following functions G is also a member of $\mathfrak{F}^p$:*

(i) $G(x,y,z)=F(y,z,x)$

(ii) $G(x,y,z)=F(g(x),y,z)$

(iii) $G(x,y,z)=g(F(x,y,z))$

(iv) $G(x,y,z)=x+y+z$.

Here, $C[R^3]=C^0[R^3]$ is the class of continuous real $(=R)$ valued functions on R^3 and $C^p[R]$ contains those with continuous kth derivatives for $k \leqslant p$. From this definition, it is easy to see that every function F in $\mathfrak{F}^p$ must have the following form:

$$F(x,y,z)=f(\phi(x)+\psi(y)+\theta(z)) \tag{1}$$

where f, ϕ, ψ, θ are in $C^p[R]$.

Suppose, now, that we are given a specific function F, say $xy+yz+zx$, and asked if this belongs to $\mathfrak{F}^p$ for some particular choice of p. While Definition 1 describes $\mathfrak{F}^p$, it does not seem directly applicable; at first glance, (1) seems far more useful, but to apply it requires us to solve a functional equation of a strange type. (See **[1]**.)

However, if F is smooth, and if we are interested in knowing if F belongs to $\mathfrak{F}^p$ when p is large enough, differentiation is the obvious tool. Indeed, if $p \geqslant 2$, every F in $\mathfrak{F}^p$ must satisfy the differential system

$$F_x F_{yz} = F_y F_{zx} = F_z F_{xy} \tag{2}$$

everywhere in R^3. This is easily verified by observing that, if F has the form (1), then the common value of each expression in (2) is $f'f''\phi'\psi'\theta'$. Upon application to the example $F(x,y,z)=xy+yz+zx$, we see that this function is not in $\mathfrak{F}^2$. However, this method, too, fails if we wish to tell whether F is a member of either of the larger classes $\mathfrak{F}^1$ or $\mathfrak{F}^0$. (See **[1]**.)

Does the nonlinear system (2) in fact characterize the class $\mathfrak{F}^2$ and, if so, is there a route that leads directly from the stated form of Definition 1 to (2)? One direction is encouraging. The set of C^2 solutions of (2) obeys each of the conditions (i)-(iv) given above, when $p=2$, showing again that every member of $\mathfrak{F}^2$ is a solution of (2). More indeed is true. Using only elementary analysis, one sees that if F is any solution of (2), then F is locally of the form (1) at all points in the complement of the set where $F_x F_y F_z = 0$. Thus, (2) is almost an equivalent characterization of the class $\mathfrak{F}^2$ but is evidentially not appropriate for $\mathfrak{F}^1$ or $\mathfrak{F}^0$.

The peculiar nature of this type of problem is illustrated by the fact that the *two*-variable version of the classes $\mathfrak{F}^p$ behaves·differently. The functions $F(x,y)$ of the form $f(\phi(x)+\psi(y))$ can be described by the same four conditions given in Definition 1 merely by suppressing "z". However, the analogue of (2) turns out to be a single *third* order (!) nonlinear differential equation

$$\{F_x F_{xyy} - F_y F_{xxy}\} F_x F_y + \{F_y{}^2 F_{xx} - F_x{}^2 F_{yy}\} F_{xy} = 0. \tag{3}$$

As before, the C^3 functions that satisfy this equation are locally of the correct form everywhere in the complement of the set where $F_x F_y = 0$ but need not have this form globally. An example is the function

$$F(x,y)=\begin{cases} \sqrt{x^8+y^8} & \text{where } x \geqslant 0, y \geqslant 0 \\ x^8+y^4 & \text{where } x \leqslant 0, y \geqslant 0 \\ x^4+y^8 & \text{where } x \geqslant 0, y \leqslant 0 \\ (x^4+y^4)^2 & \text{where } x \leqslant 0, y \leqslant 0. \end{cases}$$

Turn now to another class of functions; let $\mathcal{G}^p$ be the subclass of $C[R^3]$ consisting of those functions F of the form

$$F(x,y,z)=f(\phi(x,y)), \qquad \psi(y,z) \tag{4}$$

where f, ϕ, and ψ belong to $C^p[R^2]$. An alternative description of $\mathcal{G}^p$ can be modeled on that for $\mathfrak{F}^p$:

DEFINITION 2. *$\mathcal{G}^p$ is the smallest subclass of $C[R^3]$ with the property that, for any $F \in \mathcal{G}^p$ and $g \in C^p[R^2]$, each of the following functions is a member of $\mathcal{G}^p$:*

(i) $G(x,y,z)=F(z,y,x)$

(ii) $G(x,y,z)=F(g(x,y),y,z)$
(iii) $G(x,y,z)=g(x,z)$.

These conditions do not seem essentially different from those involved in Definition 1, so that one would expect that a differential characterization analogous to (2) ought to be possible. Unfortunately, the only one known is far from simple, involves 4th order derivatives, and is a single nonlinear partial differential equation with 55 terms! (Available from the author by request.) While this permits one to decide if a *smooth* function is locally in $\mathcal{G}^4$, it is of no help in examining membership of a specific function in any of the classes $\mathcal{G}^p$ for $p=0,1,2,3$; and, of these, criteria for membership in the largest, $\mathcal{G}^0$, is the most challenging.

In [4], Pólya and Szegő gave a simple argument to show that the trial function $F(x,y,z)=xy+yz+zx$ does not belong to $\mathcal{G}^3$ on all of R^3. The following result shows that F is not locally in $\mathcal{G}^1$ anywhere.

THEOREM. *The function $F(x,y,z)=Axy+Byz+Czx+Dy^2$ is nowhere locally in $\mathcal{G}^1$, unless $AB=CD$.*

Proof. Suppose that F belongs to $\mathcal{G}^1$ on a convex open set $\mathcal{O}$. If one of A,B,C is not zero, we can assume that $\mathcal{O}$ is disjoint from the set where $F_x=F_z=0$. Since F has the form (4) in $\mathcal{O}$, and since $\phi_1\psi_2\neq 0$ in $\mathcal{O}$, we can write

$$\begin{aligned} F_y &= (\phi_2/\phi_1)F_x+(\psi_1/\psi_2)F_z \\ &= \alpha(x,y)F_x+\beta(y,z)F_z \end{aligned} \tag{5}$$

on the set $\mathcal{O}$. With the given function F, (5) becomes

$$Ax+Bz+2Dy=(Ay+Cz)\alpha(x,y)+(By+Cx)\beta(y,z). \tag{6}$$

Upon our solving (6) for α, the result can be written as

$$\alpha(x,y)=S(y,z)1+T(y,z)x. \tag{7}$$

Since the left side is independent of z for all x and y, while 1 and x are linearly independent functions of x, it follows that S and T must each be independent of z in $\mathcal{O}$, and (7) becomes

$$\alpha(x,y)=s(y)+t(y)x. \tag{8}$$

In the same way, we solve (6) for $\beta(y,z)$, obtaining

$$\beta(y,z)=u(y)+v(y)z. \tag{9}$$

If we put (8) and (9) back into (6), we have

$$\begin{aligned} Ax+Bz+2Dy = {} & \{Ay\,s(y)+By\,u(y)\}+\{Ay\,t(y)+Cu(y)\}x \\ & +\{By\,v(y)+Cs(y)\}z \\ & +C\{t(y)+v(y)\}xz \end{aligned}$$

holding for all (x,y,z) in $\mathcal{O}$.

Accordingly, for all y in the Y projection of $\mathcal{O}$,

$$\begin{aligned} A &= Ay\,t(y)+C\,u(y) \\ B &= By\,v(y)+C\,s(y) \\ 0 &= Ct(y)+v(y) \\ 2Dy &= Ay\,s(y)+By\,u(y), \end{aligned}$$

and from these it follows that $AB=CD$. (We note that if $AB=CD$, then the given function has the required format since $F(x,y,z)=(ax+by)(cy+dz)$ and is in $\mathcal{G}^\infty$.)

This method for $\mathcal{G}^1$, which depends on an examination of the relation (5), can be applied to treat other functions F but does not seem suitable for application to the class $\mathcal{G}^0$. Other elementary approaches have been used to examine similar questions, but much remains to be discovered. For example, Manfred v. Golitschek has informed me that neither the function

$xy+yz+zx$ nor $xy^2+yz^2+zx^2$ is locally in the class $\mathcal{G}^0$; his method exploits the monotonicity of each of these functions when two of the variables are fixed. (See [2].)

References

1. R. C. Buck, Approximate complexity and functional representation, J. Math. Anal. Appl., 70 (1979) 280–298.

2. M. v. Golitschek, Remarks on functional representation, to appear in the Proc. of the Conference on Approximation Theory (Austin, January 1980).

3. G. G. Lorentz, The 13th problem of Hilbert, in Mathematical Developments Arising from Hilbert's Problems (Proc. Symp. Pure Math., vol. 28), American Mathematical Society, Providence, R.I., 1976.

4. G. Pólya and G. Szegő, Aufgaben and Lehrsätze, vol. 1, Dover, New York, 1945, part 2, #119a.

ON MODULAR GROUP RINGS, NORMAL BASES, AND FIXED POINTS

LINDSAY N. CHILDS

Department of Mathematics and Statistics, State University of New York, Albany, New York 12222

MORRIS ORZECH

Department of Mathematics, Queen's University, Kingston, Canada K7L 3N6

1. Introduction. Let G be a finite group and K/F a Galois extension of fields with group G. A *normal* basis for K/F is one of the form $\{\sigma(\alpha)|\sigma \text{ in } G\}$ for some α in K. A familiar criterion for α to generate a normal basis is that the matrix $(\sigma\tau(\alpha))$ $(\sigma, \tau$ in $G)$ have nonzero determinant (see [1, proof of Theorem 28]). However, in one situation a much simpler criterion is available:

THEOREM 1. *Let F be a field of characteristic p, K/F a Galois extension with group G whose order is p^n. The element α of K generates a normal basis iff its trace, $T\alpha$, is nonzero. Moreover, there exists α in K with $T\alpha\neq 0$.*

We shall prove this result in section 3 by exploiting a basic fact about modular group rings.

THEOREM 2. *Let F be a (not necessarily commutative) field of characteristic p, G a group of order p^n, M a nonzero left FG-module. Then there exists $m\neq 0$ in M such that $\sigma m=m$ for all σ in G.*

This theorem is a reformulation of the known result that the only irreducible FG-module is the trivial one (F with trivial G-action) (see [8, p. 19, Cor., and p. 31, Example]). The pleasant consequences of this reformulation are: it leads to proofs of basic facts about FG-modules (section 2) shorter than those we've been able to find in the literature; it and its corollaries hold when F isn't commutative; and it leads naturally to a fixed-point theorem for p-groups acting on projective space (section 4).

2. Properties of FG. We begin with a proof of Theorem 2. Since G is a p-group, there is a normal series

$$G=G_n \rhd G_{n-1} \rhd \cdots \rhd G_0=(1)$$

with G_{r+1}/G_r cyclic of order p. Let

$$M_r=\{m \text{ in } M|\sigma m=m \text{ for all } \sigma \text{ in } G_r\}.$$

M_0 is M. We shall show that $M_r\neq(0)$ implies $M_{r+1}\neq(0)$; this will prove the theorem. Let $x\neq 0$ be in M_r and let the right coset of τ in G_{r+1} generate G_{r+1}/G_r. Then τ^p is in G_r, so $\tau^p x=x$.

Define $x_m=\sum_{k=0}^{p-1} k^m\tau^k x$ for $m=0,1,\ldots$ (for $m=k=0$ interpreting k^m as 1).

Because $\tau G_r=G_r\tau$, x_m is in M_r. It is easily checked that $\tau x_0=x_0$ and that hence x_0 is in M_{r+1}.

If $x_0 \neq 0$, we are done. For general $m>0$, using $\tau^p x = x$ and $\mathrm{char}(F)=p$ we have that

$$\begin{aligned}\tau x_m &= \sum_{k=1}^{p} (k-1)^m \tau^k x\\ &= \sum_{k=0}^{p-1} (k-1)^m \tau^k x\\ &= x_m + \text{an } F\text{-linear combination of } x_0,\dots,x_{m-1}.\end{aligned}$$

If $x_i = 0$ for $i<m$, then x_m is in M_{r+1} and if also $x_m \neq 0$ we are done. We show that if $x\neq 0$, then $x_m \neq 0$ for some m:

Suppose $x_0 = \cdots = x_{p-1} = 0$. Then for any polynomial $g(X)=\sum_{m=0}^{p-1} a_m X^m$ in $F[X]$ we have

$$\sum_{k=0}^{p-1} g(k)\tau^k x = \sum_{m=0}^{p-1} a_m \sum_{k=0}^{p-1} k^m \tau^k x = \sum_{m=0}^{p-1} a_m x_m = 0.$$

But for

$$g(X) = \frac{1}{(p-1)!}\prod_{r=1}^{p-1}(r-X),$$

$g(0)=1, g(1)=\cdots=g(p-1)=0$. Hence

$$0=\sum_{k=0}^{p-1} g(k)\tau^k x = x.$$

COROLLARY. *Let F, G be as in Theorem 2. Then*

(a) *The element $T=\sum_{\sigma\in G}\sigma$ is in every nonzero left (or right) ideal of FG.*

(b) *Any irreducible left FG-module is isomorphic to the FG-module F with trivial action.*

(c) *Every element not in the augmentation ideal J of FG is a unit in FG.*

(d) $J=\mathrm{Rad}(FG)$ *(the Jacobson radical of FG). For any left FG-module M,* $\mathrm{Rad}(M) = \mathrm{Rad}(FG)M$.

Proof. (a). Let M be a left ideal of FG and $m=\sum a_\sigma \sigma$ a nonzero element of M such that $\sigma m = m$ for all σ. Then $a_\sigma = a_1$; hence $m=a_1 T$.

(b) is an immediate consequence of Theorem 2.

(c) The augmentation ideal J is the kernel of the augmentation map $\varepsilon: FG\to F$, $\varepsilon(\sum a_\sigma \sigma)=\sum a_\sigma$. It is easily verified that for α in FG, $T\alpha = \varepsilon(\alpha)T$. Hence, if α is not in J, $\varepsilon(\alpha)\neq 0$; hence $T\alpha\neq 0$; hence T is not in the left ideal $L=\{\beta\in FG \mid \beta\alpha=0\}$. By part (a) this means $L=(0)$. Thus the F-linear map from FG to FG given by multiplication by α is one-one, hence onto. Thus α is a unit in FG.

(d) That $J=\mathrm{Rad}(FG)$ is an easy consequence of (c) and of the definition of $\mathrm{Rad}(M)$ as $\cap K$, where K ranges over all kernels of FG-homomorphisms from M to irreducible FG-modules. The inclusion $\mathrm{Rad}(FG)M\subseteq \mathrm{Rad}(M)$ is a standard result and is easily proved in much greater generality than our setting. The reverse inclusion uses our hypotheses on F and G: Let $N=M/JM$. Since $1-\sigma$ is in J for all σ in G, G acts trivially on N. So N is as an FG-module simply a vector space over F. For any $m\in M, m\notin JM$, let its image in N be $\bar{m}$. There is a homomorphism $\bar{f}: N\to F$ with $\bar{f}(\bar{m})\neq 0$. $\bar{f}$ lifts to an FG-homomorphism $f: M\to F$ with kernel containing JM and with $f(m)\neq 0$. Hence m is not in $\mathrm{Rad}(M)$. Hence $\mathrm{Rad}(M)\subseteq JM$.

3. Normal bases. We now prove Theorem 1. Choose α in K. If $\{\sigma(\alpha)\mid\sigma \text{ in } G\}$ is linearly independent over F then α generates a normal basis. If not, the ideal $L=\{\beta\varepsilon FG\mid\beta\alpha=0\}$ is nonzero, hence contains T by part (a) of the corollary above. But $T\alpha=\mathrm{tr}(\alpha)$, proving the first statement of the theorem. The existence of an element of nonzero trace is a special case of the

linear independence of characters ([4, p. 209, Theorem 7]).

Theorem 1 was proved for G cyclic by Perlis ([6]). Although the criterion for general G has been noted (see [3]), it is not widely known.

4. Actions on projective space. Let F, G, M be as in Theorem 2, with $\dim_F M = d+1$. Let $\mathbf{P}$ be the set of one-dimensional F-subspaces of M; $\mathbf{P}$ may be viewed as the projective space $P^d(F)$. G permutes the elements of $\mathbf{P}$. It is clear that if $\sigma m = m$ for all σ in G then Fm is an element of $\mathbf{P}$ fixed by G. But a converse is true: if G fixes an element Fm of $\mathbf{P}$ then $\sigma m = m$ for all σ in G. For in this case

$$\sigma^{-1} m = a(\sigma) m \quad \text{for some } a(\sigma) \neq 0 \text{ in } F.$$

It follows easily that $a(\sigma\tau) = a(\sigma)a(\tau)$; hence $a(\ )$ is a homomorphism from G to F^*. But $|G| = p^n = s$ so that $a(\sigma)^s = 1$ for σ in G, implying $a(\sigma) = 1$ since $\operatorname{char}(F) = p$. This shows $\sigma m = m$ for all σ in G. We have thus shown that Theorem 2 is equivalent to the existence of a one-dimensional F-subspace of M that is fixed by G under the action induced on the set of F-subspaces of G.

Two special cases are of particular interest—F finite and F algebraically closed. Suppose $|F| = q < \infty$. Then $P^d(F)$ has $(q^d - 1)/(q-1)$ elements, so that $|\mathbf{P}| \equiv 1 \pmod p$. Each orbit of G in $\mathbf{P}$ has either 1 or p^r elements for some $r > 0$; hence some orbit has 1 element. We see that for F finite Theorem 2 is provable via a counting argument.

If F is algebraically closed, then $P^d(F)$ has the structure of an algebraic variety. Write $\mathbf{x} = (x_0 : \cdots : x_d)$ for a typical point of $P^d(F)$. To say a map σ from $P^d(F)$ to itself is a morphism of varieties means $P^d(F)$ is covered by Zariski open sets U_i such that, on each U_i, σ is given by polynomials, i.e., $\sigma(\mathbf{x}) = (H_0(\mathbf{x}) : \cdots : H_d(\mathbf{x}))$ with the H_j homogeneous of the same degree. Suppose σ is an automorphism of $P^d(F)$ as variety. It is known that then σ is induced by an F-linear automorphism of F^{d+1}; i.e., there is an invertible $(d+1)\times(d+1)$ matrix $H = (h_{ij})$ such that for all $\mathbf{x}$ in $P^d(F)$

$$\sigma(\mathbf{x}) = (H_0(\mathbf{x}) : \cdots : H_d(\mathbf{x}))$$

where

$$H_i(\mathbf{x}) = \sum_{j=0}^{d} h_{i+1,j+1} x_j.$$

(A proof of this may be found in [2, p. 151, Example 7.1.1]; a more elementary proof will appear in [5, Theorem 9.5].) Since the only elements (h_{ij}) of $\mathrm{GL}(d+1, F)$ inducing the identity on $P^d(F)$ are the scalars in F^*, this result says that the automorphism group of $P^d(F)$ is $\mathrm{PGL}(d+1, F)$, the quotient group $(\mathrm{GL}(d+1, F)/F^*$.

When F is algebraically closed, the element H of $\mathrm{GL}(d+1, F)$ that lifts σ may be found in $\mathrm{SL}(d+1, F)$, the group of matrices of determinant one (it is necessary for this only that mth roots of elements of F be in F, where $m = \operatorname{order}(\sigma)$). We thus have an exact sequence of groups

$$1 \to U \to \mathrm{SL}(d+1, F) \to \operatorname{Aut}(P^d(F)) \to 1,$$

where U is $F^* \cap \mathrm{SL}(d+1, F)$, the group of $(d+1)$st roots of unity in F^*.

Let G be a group with p^n elements, acting as a group of automorphisms of $p^d(F)$. G lifts to a subgroup S of $\mathrm{SL}(d+1, F)$ with $S/U \equiv G$. Because U and G are finite, so is S. We have an exact sequence

$$1 \to U \to S \to G \to 1.$$

U has no elements of order p^r for $r > 0$ because $\operatorname{char}(F) = p$. Hence the orders of U and G are relatively prime and S must be a semi-direct product of U by G, by the Schur-Zassenhaus Lemma ([8, Theorem 7.17 or Theorem 7.19]). Hence there is a group T in S, with $T \cong G$, acting

as a group of automorphisms on F^{d+1} and lifting the action of G on $P^d(F)$. By Theorem 2, there is a fixed point for T in F^{d+1}; hence there is one for G in $P^d(F)$. We have proved:

THEOREM 3. *Let F be an algebraically closed field of characteristic p, G a group with p^n elements acting as F-automorphisms of $P^d(F)$. Then there is a point of $P^d(F)$ left fixed by G.*

5. A characterization of p-groups. A. R. Magid and I. Reiner have pointed out to us that there is a converse to Theorem 2.

THEOREM 4. *Let F be a field of characteristic $p \neq 0$, G a finite group. If every nonzero left FG-module has a fixed point, then G is a p-group.*

Proof. The number r of nonisomorphic irreducible FG-modules is the number of conjugacy classes of p'-elements of G, i.e., of elements of G whose order is relatively prime to p (see [**8**, p. 17, Theorem 1.5 and p. 31, Example]). If G is not a p-group, then $r>1$. But F is the only irreducible FG-module with a fixed point.

We wish to thank L. McCulloh and I. Hughes for helpful comments.

Research for this paper was supported in part by NSF grant MCS 76-03946 and in part by NSERC grant A7202.

References

1. E. Artin, Galois Theory, Notre Dame Mathematical Lectures Number 2, University of Notre Dame Press, 1959.

2. R. Hartshorne, Algebraic Geometry, Springer-Verlag, New York, 1977.

3. H. F. Kreimer, Normal bases for Galois p-extensions of rings, Notices Amer. Math. Soc., 24 (1977) A-268.

4. S. Lang, Algebra, Addison-Wesley, Reading, Mass., 1965.

5. G. Orzech and M. Orzech, Plane Algebraic Curves, Marcel Dekker, New York, to appear.

6. S. Perlis, Normal bases of cyclic fields of prime-power degree, Duke Math. J. (1942) 507–517.

7. J. J. Rotman, The Theory of Groups, Allyn and Bacon, Boston, 1973.

8. B. M. Puttaswamaiah and J. D. Dixon, Modular Representations of Finite Groups, Academic Press, New York, 1977.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4.*

SEQUENCES OF POLYHEDRA

L. FEJES TÓTH

Hungarian Academy of Sciences, Budapest V., Realtanoda u. 13-15, Hungary

In the Euclidean plane, let P_0 be a convex polygon, P_1 the polygon spanned by the side-midpoints of P_0, P_2 the polygon spanned by the side-midpoints of P_1, and so on. Cadwell

[1], [2] proved that in the sequence $P_0, P_1, P_2, \ldots$ the polygons will be more and more similar to the affine image of a regular polygon. More precisely, there is a set of affine transformations $a_0, a_1, a_2, \ldots$ such that the limiting figure of the set $a_0P_0, a_1P_1, a_2P_2, \ldots$ is a regular polygon.

This nice result remained almost unnoticed until it was partly rediscovered by this author [3], who gave a geometrical proof for polygons with less than seven sides and suggested several analogous problems. The paper [3] started a set of investigations dealing with various generalizations and variants of the original problem, giving new proofs and rediscovering old ones [4]–[13]. The sequence of polygons $P_0, P_1, \ldots$ was studied, independently of Cadwell, also in the papers [14] and [15].

In 3-space, problems of another type present themselves. Given a convex polyhedron P we can consider, for instance, the convex hull of the centroids of the faces of P. Starting with an affinely regular solid, iteration of this process yields smaller and smaller copies of the solid and its dual. But a small distortion of the starting solid may cause a radical change in the sequence of polyhedra. Starting, for instance, with a polyhedron topologically isomorphic to the regular octahedron, in general we will obtain, in the first step, a polyhedron with 8 vertices and $2 \cdot 8 - 4 = 12$ triangular faces; in the second step, a polyhedron with 12 vertices and $2 \cdot 12 - 4 = 20$ triangular faces, and so on. Thus this process has not the effect of making the shape of the polyhedron more and more regular, only more and more complicated. Omitting the problems arising in connection with this process, we consider another analogue of the above sequence of polygons.

Let P_0 be a convex polyhedron, and let P_i $(i = 1, 2, \ldots)$ be the convex hull of the *edge*-midpoints of P_{i-1}. Since $P_i \subset P_{i-1}$, and P_i contains the centroids of the faces of all of its predecessors $P_{i-1}, \ldots, P_0$, the limiting figure $\lim_{i \to \infty} P_i$ is a nondegenerate convex body which we call the **kernel** of P_0. From among the various problems concerning the kernel of a convex polyhedron, we call attention to the following ones:

1. Prove or disprove the conjecture that among the convex polyhedra of equal volume the kernel of the tetrahedron has the least possible volume.

2. What can be said about the structure of the boundary of the kernel? Is the kernel of any convex polyhedron smooth?

References

1. J. H. Cadwell, A property of linear cyclic transformations, Math. Gaz., 37 (1953) 85–89.
2. J. H. Cadwell, Nested polygons, Topics in Recreational Mathematics, Cambridge University Press, 1966, Ch. 3, pp. 22–31.
3. L. Fejes Tóth, Processes of iteration concerning polygons (Hungarian), Mat. Lapok, 20 (1969) 15–23.
4. F. Kárteszi, A pair of affinely regular pentagons generated from a convex pentagon (Hungarian), Mat. Lapok, 20 (1969) 7–13.
5. G. Korchmáros, A process of iteration leading to affinely regular polygons (Hungarian), Mat. Lapok, 20 (1969) 405–411.
6. M. Ruda, On a conjecture of L. Fejes Tóth (Hungarian), Magyar. Tud. Akad. Mat. Fiz. Oszt. Közl., 19 (1970) 375–381.
7. H. Reichardt, Bestätigung einer Vermutung von Fejes Tóth, Rev. Roumaine Math. Pures Appl., 15 (1970) 1513–1518.
8. G. Lükő, Certain sequences of inscribed polygons, Period. Math. Hungar., 3 (1973) 255–260.
9. A. Szép, Linear iterations of polygons (Hungarian), Mat. Lapok, 21 (1970) 363–368.
10. M. Ruda, A theorem concerning bisecting-polygons (Hungarian), Magyar Tud. Akad. Mat. Fiz. Oszt. Közl., 22 (1973) 201–213.
11. G. Bruckner, Über n-Ecke, Studia Sci. Math. Hungar., 7 (1972) 429–435.
12. G. Fejes Tóth, Processes of iteration leading to regular polygons (Hungarian), Mat. Lapok, 23 (1972) 135–141.
13. M. Bourdeau and S. Dubuc, L'itération de Fejes Tóth sur un Polygone, J. Geometry, 6 (1975) 65–75.
14. E. R. Berlekamp, E. N. Gilbert, and F. W. Sinden, A polygon problem, this MONTHLY, 72 (1965) 233–241.
15. R. J. Clarke, Sequences of polygons, Math. Mag., 52 (1979) 102–105.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

Send all proposed problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions should be sent to the addresses given at the head of each problem set.

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Isoperimetric Problem in a Disk

S 19 [1979, 702]. *Proposed by Harley Flanders, Florida Atlantic University.*

Let C be a smooth simple arc inside the unit disk, except for its endpoints, which are on the boundary. How long must C be if it cuts off one-third of the disk's area? Generalize.

Solution by Jordi Dou, Barcelona, Spain. Using Calculus of Variations, one sees that the arc C of minimal length must be a circular arc normal to the boundary of the disk at the endpoints A and B of C. Let O be the center of the disk and $\measuredangle AOB=2\theta$. Then the given area condition becomes

$$\pi/3=\theta+[(\pi/2)-\theta]\tan^2\theta-\tan\theta.$$

The approximate solution of this trancendental equation is $\theta\approx1.175$ and the length of C is $(\pi-2\theta)\tan\theta\approx1.89$.

Also solved by Michael Goldberg, Doug Hensley, Thomas Hughes, and the proposer.

Note. Dou, Goldberg, and Hughes dealt with the generalization to an arbitrary ratio for the areas cut out of the disk by C, while Hensley stated that the analogue on the minimal surface separating a sphere into regions whose volumes are in a given ratio is dealt with in J. Bokowski and E. Sperner, *Zerlegung konvexer Körper durch minimale Trennflächen*, preprint no. 236, Sonderforschungbereich 72, Approximation und Optimisierung, Universität Bonn.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by June 30, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2869. *Proposed by Desmond MacHale, Cork, Ireland.*

Let G be a group with center $Z(G)$ such that $[G\colon Z(G)]$ is finite. Prove that if H is a subgroup of G then $[G\colon Z(G)]\geqslant[H\colon Z(H)]$. Find necessary and sufficient conditions for equality and show that, in the case of equality, $G/Z(G)\simeq H/Z(H)$ and $G'=H'$ where prime denotes the commutator subgroup.

E 2870. *Proposed by Gideon Schwarz, Hebrew University, Jerusalem.*

Find all positive integers n such that $3^n + 4^n + \cdots + (n+2)^n = (n+3)^n$ (Example: $3^2 + 4^2 = 5^2$).

E 2871. *Proposed by W. G. Leavitt, University of Nebraska.*

Let n be a positive integer relatively prime to 10. Call n a nines number if, for every h relatively prime to n, the decimal expansion of h/n has even period, and the sum of the two half-periods is 999.... Show that n is a nines number if and only if $10^k \equiv -1 \pmod n$ for some k.

E 2872. *Proposed by J. G. Mauldon, Amherst College.*

Find five different triples of positive integers with the same sum and the same product.

SOLUTIONS OF ELEMENTARY PROBLEMS

Traversing Linearly Ordered Sets

E 2608* [1976, 567]. *Proposed by Judith Q. Longyear, Wayne State University.*

A child is riding in a train n cars long and wishes to go exploring. An exploration may be described by listing in order the cars traversed; each exploration must end in the same car in which it began. How many explorations of length k can it make?

Suppose we regard the exploration $(e_1, e_2, \dots, e_k)$ to be equivalent with all explorations $(e_r, e_{r+1}, \dots, e_k, e_1, \dots, e_{r-1})$ $(r = 2, \dots, k)$. How many nonequivalent explorations can it make?

Solution by Allen J. Schwenk, U.S. Naval Academy. The first question is just the number of closed k-walks in the path P_n. This is the trace of the kth power of the adjacency matrix A for the path. The eigenvalues of A are well known; they are $\lambda_j = 2\cos \pi j/(n+1)$. Thus $\operatorname{Tr} A^k = \sum_{j=1}^n \lambda_j^k = 2^k \sum_{j=1}^n \cos^k \pi j/(n+1)$. Of course, when k is odd this formula gives 0, since $\lambda_j = -\lambda_{n+1-j}$ according to the pairing theorem for bigraphs.

The second part of the question is an elementary application of Burnside's Lemma. For any graph G, the number of nonequivalent closed k-walks is

$$\frac{1}{k}\sum_{d|k}\phi\left(\frac{k}{d}\right)\sum_{j=1}^{n}\lambda_j^d$$

where ϕ is the Euler phi function. For the path P_n this yields

$$\frac{1}{k}\sum_{d|k}\phi\left(\frac{k}{d}\right)2^d\sum_{j=1}^{n}\cos^d \pi j/(n+1).$$

Again, the terms with d odd provide a net contribution of 0 to the sum.

A similar formula applies if it is postulated that the reversed walk $(e_k, e_{k-1}, \dots, e_1)$ is also equivalent to $(e_1, \dots, e_k)$. [The cyclic group of order k is replaced by the dihedral group of order $2k$.]

All these results for arbitrary graphs, and for open walks as well as closed, appear in greater detail in my article *The spectral approach to determining the number of walks in a graph*, coauthored with Frank Harary, Pacific J. Math., 80 (1979) 443–449, and also in my Ph.D. dissertation, University of Michigan, 1973.

An Easy Counterexample

E 2797 [1979, 785]. *Proposed by Barry J. Powell, Kirkland, Washington.*

Show that there are infinitely many primes p such that for any pair of coprime odd positive

integers x and y with no two of p, x, y, congruent modulo p, the exponent of p in $x^{p-1}-y^{p-1}$ is odd. [The exponent of p in m is the largest integer e such that $p^e|m$.]

Counterexample by Robert E. Shafer, Berkeley, California. Take $x=p^2+2$, $y=p^2-2$. Then $p^3\nmid x^{p-1}-y^{p-1}$, $p^2|x^{p-1}-y^{p-1}$.

Counterexamples were also found by L. L. Foster and University of South Alabama Problem Group.

Integrals of Trigonometric Functions

E 2803 [1979, 864]. *Proposed by L. R. Shenton, Frank Bowman, and H. K. Lam, University of Georgia.*

Prove

(a) $\int_0^{\pi/4} g(\theta)d\theta=\pi^2/24$,

(b) $\int_0^{\pi/6} g(\theta)d\theta=\pi^2/32$,

where $g(\theta)=\arctan[(\cos 2\theta)/(2\cos^2\theta)]^{1/2}$.

Solution by St. Olaf College Problem Group. To evaluate (a), make the substitution $\tan\theta=\sin x$; this transforms the integral into the form

$$G(\alpha)=\int_0^{\pi/2}\arctan(\alpha\cos x)\frac{\cos x}{1+\sin^2 x}dx,$$

where $\alpha=1/\sqrt{2}$. Differentiating this relation with respect to α and making the substitution $t=\tan x$ gives the value

$$G'(\alpha)=\frac{\pi}{2}\left[\sqrt{2}/(1+2\alpha^2)-1/(1+2\alpha^2)\sqrt{1+\alpha^2}\right].$$

Integrating, and noting that $G(0)=0$, gives the result. To evaluate (b) consider the integral

$$K(\alpha)=\int_0^{\delta}\arctan(\alpha\cos x)\frac{\cos x}{1+\sin^2 x}dx,$$

where $\delta=\arcsin 1/\sqrt{3}$. Differentiating with respect to α allows the integrals with respect to x to be evaluated. Integrating back with respect to α gives

$$K(1/\sqrt{2})=\frac{\pi}{4}\int_0^{1/\sqrt{2}}(1+2\alpha^2)^{-1}\sqrt{2}\,d\alpha-\int_0^{1/\sqrt{2}}\frac{\arctan\dfrac{1}{\sqrt{2}\sqrt{1+\alpha^2}}}{(1+2\alpha^2)\sqrt{1+\alpha^2}}d\alpha.$$

The first integral is equal to $\pi^2/16$. In the second, make the substitution $\tan\beta=\alpha$; this results in the relation $K(1/\sqrt{2})=\pi^2/16-K(1/\sqrt{2})$ and the result now follows.

Also solved by Paul F. Byrd, Otto G. Ruehr, William V. Webb, and the proposers.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor R. C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by June 30, 1981. The solver's full post-office address should be on each sheet.

6330. *Proposed by Ko-Wei Lih, Academia Sinica, Taiwan*

Characterize all $n\times n$ complex matrices A for which relation $\mathrm{per}(AX)=\mathrm{per}(A)\mathrm{per}(X)$ holds for every complex $n\times n$ matrix X. (Here $\mathrm{per}(B)$ denotes the permanent of B.) Does the same characterization hold when the complex field is replaced by an arbitrary field?

6331. *Proposed by D. Hensley, Texas A&M University.*

Suppose B is a centrally symmetric bounded convex body of uniform density half that of water and it floats with neutral buoyancy in any orientation. Prove that B must be a ball.

6332. *Proposed by Petter E. Bjørstad, Stanford University, and Henry E. Fettis, Mountain View, California.*

In analyzing an accelerated numerical integration method for the biharmonic equation, the finite sum

$$S_N = \sum_{k=1}^{N-1} \sin^2(k\pi/N) \Big/ \left[1+a^2-2a\cos(k\pi/N)\right]^2$$

arises ($0 \leqslant a^2 < 1$). Find a closed algebraic expression for S_N that exhibits its asymptotic behavior in terms of a and N.

6333. *Proposed by M. J. Pelling, University of Malaya.*

Let N_* denote the set of infinite integer sequences $n_* = (n_1, n_2, \ldots)$, all components being nonnegative, partially ordered according to: $m_* \leqslant n_*$ if $\forall_i[m_i \leqslant n_i]$. Show that the following question is undecidable. Does N_* contain an uncountable subset X_* such that, for every member k_* of N_*, the set

$$\{n_* | n_* \in X_* \text{ and } n_* \leqslant k_*\}$$

is countable?

SOLUTIONS OF ADVANCED PROBLEMS

Increasing Polynomials in an Ordered Field

5861 [1972, 667; 1975, 767]. *Proposed by Michael Slater, University of Bristol, England.*

Let F be an ordered field.

(a) If $p \in F[x]$; $a, b \in F$, $a < b$, and $p'(x) > 0$ for $a \leqslant x \leqslant b$, does it follow that $p(a) < p(b)$?

(b) If Rolle's theorem holds in F, does it follow that F is real-closed?

Solution to part (b) by M. J. Pelling, University of Malaya, Kuala Lumpur. The answer is negative as will be shown by constructing an ordered field F which is not real-closed but for which Rolle's theorem holds. We shall use the nomenclature from the solution to part (a) [1975, 767] of finite and infinitesimal elements in an ordered field and of the standard part map h.

Let $Q(z)$ denote the field of rational functions in a complex variable z over Q. By an algebraic function will be meant any complex function $f(z)$ defined and analytic in a domain D of the form $0 < |z| < r_f$, $0 \leqslant \arg(z) < 2\pi$ cut along the positive x-axis, and which satisfies a polynomial equation over $Q(z)$ so that there are rational functions $A_0(z), A_1(z), \ldots, A_n(z) \in Q(z)$ with

$$A_0 f^n + A_1 f^{n-1} + \cdots + A_n = 0 \quad \text{in } D.$$

The set of algebraic functions forms a field K and, if the elements of $Q(z)$ are also regarded as algebraic functions, then $Q(z)$ is a subfield of K and K is the algebraic closure of $Q(z)$.

Any member of K may be represented uniquely in the form

$$f(z) = \alpha z^{m/n}\left(1 + \sum_{i=1}^{\infty} \alpha_i z^{i/n}\right), \qquad 0 < |z| < r_f, \qquad 0 \leqslant \arg(z) < 2\pi \tag{1}$$

where m, n are integers, $n \geqslant 1$, the α and α_i are algebraic numbers over Q, and where if $z = re^{i\theta}$, $0 \leqslant \theta < 2\pi$, then $z^{1/n}$ is defined (single valued) as $r^{1/n}e^{i\theta/n}$. In the representation (1) it is assumed

that no common factor can be canceled out of n, m and all i for $\alpha_i \neq 0$ so that n is minimal, and in the usual terminology the algebraic function $f(z)$ has a branch point of order n at $z=0$. The integer $n=n(f)$ will be called the order of f.

The set L of elements of K for which all α and α_i are real is a subfield of K and can be ordered by defining $f>0 \Leftrightarrow \alpha>0$, which also induces an ordering of $Q(z) \subseteq L$. With these orderings, since $K=L(i)$, L is the real-closure of $Q(z)$. In the representation (1) an element $f \in L$ is finite iff $m \geqslant 0$ and is infinitesimal if $m>0$; if finite the standard part $h(f)$ is given by the real algebraic number $h(f)=f(0)$, allowing the obvious extension of the domain of f to include 0. If f is not finite then $m<0$ and we write $h(f)=f(0)=\infty$.

The ordered field F is defined as $\{f \mid f \in L \;\&\; n(f) \text{ odd}\}$ which the reader should check is really a field. Then F is not real-closed since, e.g., $z^{1/2} \in L-F$, and it remains to show that Rolle's theorem is valid in F. By linear changes of variable if necessary it suffices to prove that if $p(x)$ is a polynomial over F such that $p(0)=p(1)=0$ then $p'(x)$ has a root in the open interval $(0,1)$ in F. Excluding the trivial case $p \equiv 0$ and dividing by the coefficient of largest absolute value if necessary we may assume that all coefficients of p are finite and not all are infinitesimal. Let $q \in R[x]$ be the real polynomial resulting from p under h so that q' results from p' under h.

Since $p \not\equiv 0$ and $p(0)=0$ and not all coefficients of p are infinitesimal, $q' \not\equiv 0$. If $q' \geqslant 0$ in the real open interval $(0,1) \subseteq R$ then

$$q(1)-q(0)=\int_0^1 q'(x)\,dx>0 \Rightarrow p(1)>p(0).$$

A similar contradiction results on the assumption that $q' \leqslant 0$ in $(0,1) \subseteq R$, whence it follows that q' must have a root θ, a real algebraic number with $0<\theta<1$, of odd multiplicity $2j+1$, say.

Let $p'(x)=f_0(z)x^k+f_1(z)x^{k-1}+\cdots+f_k(z)$ so that

$$q'(x)=f_0(0)x^k+f_1(0)x^{k-1}+\cdots+f_k(0)$$

and if $g_1(z),\ldots,g_k(z)$ are the k roots in K of $p'(x)=0$, multiple roots counted accordingly, then $g_1(0),\ldots,g_k(0)$ are the roots of $q'(x)=0$, provided we adopt the convention that if the first r coefficients of $q'(x)$ vanish then q' has ∞ as an r-fold root. Let $g_1, g_2, \ldots g_{2j+1}$ be the $2j+1$ roots of $p'(x)$ such that $g_i(0)=\theta$. It will be shown that at least one of these roots falls in F.

If $n=\text{l.c.m.}\{n(f_i) \mid 0 \leqslant i \leqslant k\}$, which is odd since all $f_i \in F$, then introducing the complex variable ζ by $z=\zeta^n$ the coefficients $f_i(z)=f_i(\zeta^n)$ of $p'(x)$ become functions of ζ analytic in some disc $|\zeta|<r$. The roots $g_i(z)=g_i(\zeta^n)$ may also be expressed in terms of ζ, although these may only be analytic in a cut disc.

If $g=g_i$, $1 \leqslant i \leqslant 2j+1$, is a real root (that is, in L) of $p'(x)$, then g admits a representation

$$g=\theta+\sum_1^\infty a_m \zeta^{m/p}, \qquad a_m \text{ real},$$

where g, as a function of ζ, has a branch point of order p at $\zeta=0$. Since the coefficients f_i are invariant under $\zeta^{1/p} \to e^{2\pi ik/p}\zeta^{1/p}$, $0 \leqslant k<p$, it follows that

$$g^{(k)}=\theta+\sum_1^\infty a_m e^{2\pi ikm/p}\zeta^{m/p}$$

is also a root of $p'(x)$ for each $k, 0 \leqslant k<p$. If p is odd then none of the additional roots is real, but if p is even then $k=p/2$ gives one additional real root $\theta+\sum_1^\infty(-1)^m a_m \zeta^{m/p}$.

Any root g_i, $1 \leqslant i \leqslant 2j+1$, which cannot be obtained in this way must be a complex root of the form $g_i(z)=a(z) \pm i\sqrt{b(z)}$ where $a, b \in L$, $b>0$, $b(0)=0$, $a(0)=\theta$, so that the remaining roots fall into complex conjugate pairs, say q pairs in all.

It follows that the set of roots $g_1,\ldots,g_{2j+1}$ divide into disjoint groups of $p_1, p_2, \ldots$ roots, together with a group of $2q$ complex roots, such that a group of p_i roots contains 1 or 2 real roots according as p_i is odd or even. But $2j+1=2q+\sum p_i$, so that at least one $p_i=p$ must be odd. If $g=\theta+\sum_1^\infty a_m \zeta^{m/p}$ is the corresponding real root in the ith group then

$$g(z)=\theta+\sum_{1}^{\infty}a_m z^{m/np}$$

is a real root g of $p'(x)$ such that $0<g<1$ and $n(g)|np$, so that $n(g)$ is odd and $g\in F$. This proves Rolle's theorem for F.

Although an ordered field F satisfying Rolle's theorem need not be real-closed, one can show that any polynomial over F of odd degree has a root in F.

A substantial partial solution, with some generalizations, was received from F. Rowbottom.

Nesting Regular n-gons

6062* [1975, 1016; 1977, 578]. *Proposed by B. H. Voorhees, University of Alberta.*

Consider an infinite sequence of regular n-gons such that each $(n+1)$-gon is contained within the preceding n-gon and is of maximal area consistent with this constraint. Take the first element of this sequence as an equilateral triangle having unit area. Is the limit of this sequence a point or a circle? If it is a circle, determine its area.

A complete solution is given by B. A. Troesch, *Optimization of nested polygons*, J. Optimization Theory and Applications, 31 (1980) No. 2, to appear. (See also Abstract 78T-D9, Notices A.M.S. 25 (1978) A-359.) The author determines the size and position of the n-gons (which are off center), and determines the ratio of the radius of the limit circle to that of the incircle of the initial equilateral triangle as an infinite product, with value 0.341473.

A substantial partial solution was obtained by D. J. Daley (Australia).

An Isoperimetric Problem

6076* [1976, 141]. *Proposed by Robert L. Anderson, Spartanburg, South Carolina.*

Given n real numbers $p_1,p_2,\ldots,p_n$, find a continuous function $x(t)$ with piece-wise continuous derivative $x'(t)$ on $[0,n]$ such that $x(t)$ minimizes

$$L(x)=\int_0^n\sqrt{1+[x'(t)]^2}\,dt$$

subject to the n constraints

$$\int_{i-1}^{i}x(t)\,dt=p_i,\qquad i=1,2,\ldots,n.$$

Is the solution unique?

Solution by M. J. Pelling, University of Malaya, Kuala Lumpur. It will be shown if a solution exists then it is unique but also that a solution may not exist at all. By well-known isoperimetric results on minimal arcs connecting two points, or joining one point to a line, and bounding a prescribed area, a solution $x(t)$ must comprise a circular arc G_i in each interval $[i-1,i]$, $1\leqslant i\leqslant n$, with the end arcs G_1 and G_n meeting the respective ordinates $t=0$ and $t=n$ at right angles.

It is also true here that the arcs G_i and G_{i+1} must join smoothly at $(i,x(i))$, in that $x'(i)$ exists, possibly $\pm\infty$ valued, and is continuous at $t=i$. This is easily proved by mechanical considerations if we imagine the curve $x(t)$ as an inextensible impermeable string passing through an incompressible 2-dimensional fluid F_i filling each rectangle $i-1\leqslant t\leqslant i$, $-K\leqslant x\leqslant K$ for some fixed large K, where the area of fluid F_i in the region $i-1\leqslant t\leqslant i$, $-K\leqslant x\leqslant x(t)$ is $K+p_i$. The fluids F_i are separated by the ordinates $t=i$, but we suppose the string to pass freely through each ordinate $t=i$ via a small ring R_i which slides frictionlessly along the ordinate. If the string is pulled tight at the ends then, since $x(t)$ already minimises $L(x)$, the string does not alter its position, which means the rings R_i are in equilibrium and the gradient of the string on the left of R_i is the same as the gradient on the right. In other words $x'(i)$ exists and G_i,G_{i+1} join smoothly.

We now prove that $x(t)$ is unique, if it exists, by showing that there can be at most one function $x(t)$ with the following properties.

1) In each interval $[i-1,i]$, $1 \leqslant i \leqslant n$, $x(t)$ is a circular arc G_i and the arcs G_i, G_{i+1} join smoothly at $(i, x(i))$;
2) $x'(0) = x'(n) = 0$, that is the arcs G_1, G_n meet the respective ordinates $t = 0, t = n$ at right angles;
3) $\int_{i-1}^{i} x(t)\,dt = p_i$.

Suppose $\bar{x}(t)$ were another function on $[0,n]$ with the same properties and corresponding arcs $\bar{G}_i$. We could not have $\bar{x}(0) = x(0)$, since by (2) and (3) for $i = 1$ it would follow $\bar{G}_1 = G_1, \bar{x}(1) = x(1), \bar{x}'(1) = x'(1)$, whence by (3) for $i = 2$ it would follow $\bar{G}_2 = G_2$ and iterating this argument $\bar{x}(t) = x(t)$ for $0 \leqslant t \leqslant n$. Assume therefore without loss of generality that $\bar{x}(0) > x(0)$.

LEMMA. *If $y(t), z(t)$, $0 \leqslant t \leqslant 1$ are two circular arcs such that $y(0) > z(0)$, $y'(0) \geqslant z'(0)$, and $\int_0^1 y(t)\,dt = \int_0^1 z(t)\,dt$, then $y(1) < z(1), y'(1) < z'(1)$.*

Proof of lemma. The hypotheses $y(0) > z(0), y'(0) \geqslant z'(0)$ imply that the two arcs cannot meet twice in [0, 1] (draw a diagram) and the hypothesis $\int_0^1 y(t)\,dt = \int_0^1 z(t)\,dt$ then implies that they cross exactly once in (0, 1), so that $y(1) < z(1)$. Examination of cases shows that the configuration is only possible with $y'(1) < z'(1)$.

Since $\bar{x}'(0) = x'(0)$ it follows from the lemma that $x(1) > \bar{x}(1), x'(1) > \bar{x}'(1)$, and applying the lemma again that $\bar{x}(2) > x(2), \bar{x}'(2) > x'(2)$ and so on up to $t = n$, at which point the requirement $\bar{x}'(n) = x'(n)$ of (2) will be contradicted. This completes the proof that a solution $x(t)$ must be unique.

Finally, it follows from (1) and (2) when $n = 2$ that $|p_1 - p_2| \leqslant \pi/2$, so there can be no solution if $|p_1 - p_2| > \pi/2$. Note, however, that the mechanical method used earlier also provides a limiting form of solu'ion in such cases as a curve with continuously varying tangent which is a circular arc in each interval $(i-1, i)$ and meets each ordinate $t = i$ in either a point or a segment.

Distribution of Inner Product of Two Random Vectors

6207 [1978, 282]. *Proposed by Ignacy I. Kotlarski, Oklahoma State University.*

Let X, Y be two independent $(2n+2)$-dimensional normal random vectors with means 0 and positive definite variance covariance matrices C, C^{-1} respectively $(n=0,1,\ldots)$. Find the distribution of their inner product $Z = X \cdot Y$.

Solution by G. S. Rogers, New Mexico State University. Let $X[Y]$ be a p dimensional normal random variable (rv) with mean zero and variance-covariance matrix $C[D]$. The characteristic function (cf) of $Z = X'Y$ when X and Y are independent is $E[\exp itX'Y] = E[E[\exp itX'Y|X]] = E[\exp - t^2X'DX/2] = |I + t^2CD|^{-1/2} = \prod_{j=1}^{p}(1+t^2\lambda_j)^{-1/2}$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of CD.

When $\lambda_j > 0$, $(1+t^2\lambda_j)^{-1/2}$ is the cf of $G_j - H_j$, where G_j, H_j are independent and identically distributed (iid) gamma rvs. When all $\lambda_j > 0$, Z is distributed as the sum of the independent $G_1 - H_1, \ldots, G_p - H_p$.

When all $\lambda_j = 1$ $(D = C^{-1})$, $(1+t^2)^{-p/2} = (1+it)^{-p/2}(1-it)^{-p/2}$ and Z is distributed as the difference of two iid gamma rvs. For $w > 0$, $P(Z \leqslant w) = \int_0^\infty (\int_0^{y+w} x^{p/2-1}e^{-x}\,dx) y^{p/2-1}e^{-y}\,dy \div (\Gamma(p/2))^2$ and the corresponding density is $h(w) = \int_0^\infty (y+w)^{p/2-1}e^{-w-2y}y^{p/2-1}\,dy \div (\Gamma(p/2))^2$. For $w < 0$, the density is $h(-w)$.

Also solved by Theodore S. Bolis, Paul S. Bruckman, L. E. Clarke (England), Richard A. Groeneveld, Dave Joyner & Fred Roush, Franklin Kemp, O. P. Lossers (Netherlands), R. M. Norton, Lajos Takács, and the proposer.

When Does $AB=C$ Imply $BA=D$?

6251 [1979, 60]. *Proposed by William P. Wardlaw, U.S. Naval Academy.*

Let m and n be positive integers. What pairs of matrices C and D, over any field K, have the property that if A is an $m\times n$ matrix over K and B is an $n\times m$ matrix over K such that $AB=C$ then $BA=D$?

Solution by J. T. Arnold, Virginia P.I., Duane Broline, Auburn University, James G. Mauldon, Amherst College, and O.P. Lossers, Eindhoven U.T., Netherlands (independently). The context can be enlarged to a division ring. The conditions are, either

(i) rank $C>n$, or

(ii) $n=1$, $D=0$, $C=0$, or

(iii) rank $C=n$, and for some λ in the center of K, $D=\lambda I$, $C^2=\lambda C$. (If K is commutative and if $\lambda\neq 0$, then $C\simeq$ diag$[D,0]$.) Rank C is understood to mean the minimum value of r such that $C=(R_{m\times r})(S_{r\times m})$; r is the left row-rank (and the right column rank).

Condition (i) is a vacuous case. For condition (ii), it is easily verified that, if $n=1$,

$$\{AB=C\Rightarrow BA=D \text{ and } C=0\}\Leftrightarrow\{AB=C\Rightarrow BA=D \text{ and } D=0\}.$$

We pass to (iii).

Suppose $AB=C$. Let T be any invertible $m\times n$ matrix. Set $A_1=AT$, $B_1=T^{-1}B$. Then $A_1B_1=C$. If also $B_1A_1=D=BA$, then $T^{-1}DT=D$, so that D is commutative with every invertible matrix. Thus for some λ in the center of K, $D=\lambda I_n$. (The internal situation is different for $\lambda\neq 0, \lambda=0$)

If $AB=C$, $BA=D$, then also $ABAB=C^2=A\,DB=(\lambda I_n)AB=C$. To show that, if $n>1$, then rank $C=n$, suppose $C=(P_{m\times r})(Q_{r\times m})$, $r<n$. Take $A=[P,X]$, $B=[Q^*,0]^*$. Then $AB=C$, but $BA=[QP,QX;\ 0,0]$. Now QX varies with X unless $Q=0$. Thus $C=0$. Since $n>1$, the condition $AB=0_{m\times m}$ can be satisfied nontrivially by taking $A=[X,0]$, $B=[0,Y^*]^*$. But then, the product BA is not constant, since YX varies.

To establish the converse in case $n>1$, suppose $D=\lambda I$, $C^2=\lambda C$. Then $0=ABAB-I_nAB=A(BA-I_n)B$. Note that $A[B]$ has a left [right] inverse, since it is of rank n. Thus $BA-\lambda I_n=0$.

If K is commutative and if $\lambda=0$, i.e., if $D=0$, then C is similar to diag$[G,0]$, where G is the direct sum of n summands $\begin{bmatrix}0,1\\0,0\end{bmatrix}$. Thus $m\geqslant 2n$ in this case.

If $D\neq 0$, Mauldon asks whether C is always similar to diag$[D,0]$ in the noncommutative case.

Associativity in a Ring

6263 [1979, 226]. *Proposed by David Pokrass, Lanier Business Products, Atlanta, Georgia.*

In a simple nonassociative ring R, let

$$(a,b,c)=(ab)c-a(bc),\qquad [a,b]=ab-ba,\qquad a\circ b=ab+ba.$$

If R satisfies the identity $w\circ(x,y,z)=0$ and has no elements of additive order 2, show that R is either associative or anticommutative, i.e., R satisfies either $(x,y,z)=0$ or $x\circ y=0$ identically.

Solution by the proposer. We will assume that R is a simple ring satisfying the identity $a\circ(b,c,d)=0$. We also assume $2x=0$ implies $x=0$. We show that R is either associative or anti-commutative.

In any ring one has the following identity:

$$[x,a\circ b]-a\circ[x,b]-b\circ[x,a]=(a,x,b)+(b,x,a)-(x,a,b)$$
$$-(b,a,x)-(a,b,x)-(x,b,a). \qquad (1)$$

Let us define $T=\{t\in R\colon R\circ t=0\}$. Then T is an additive subgroup, and $(R,R,R)\subset T$. If we assume $a\in(R,R,R)$ in (1) we get $b\circ[x,a]\in T$. Since $a\in T$ implies $[x,a]=2xa$, while R

contains no element of additive order 2, we conclude that

$$R \circ R(R,R,R) \subset T. \tag{2}$$

In any ring $J=(R,R,R)+R(R,R,R)$ is an ideal. Relation (2) implies $R \circ J \subset T$. Since R is simple, either $J=0$ or $J=R$. If $J=0$, then R is associative; so we may assume $J=R$. We now have $R \circ R \subset T$. Note this implies $xy \equiv -yx \bmod T$. We also have $(xy)z=(x,y,z)+x(yz) \equiv x(yz) \bmod T$. Therefore $x(yz) \equiv (xy)z \equiv -z(xy) \equiv -(zx)\,y \equiv y(zx) \equiv (yz)x \equiv -x(yz)$. It follows that $2x(yz) \in T$, whence, under our assumptions, $R^3 \circ R=0$. Since R is simple and not associative, $R=R^3$ and we have $R \circ R=0$. That is, R is anti-commutative.

Note. A. Thedy [Proc. A.M.S., 29 (1971) 250–254] has shown that a simple ring satisfying $[a,(b,c,d)]=0$ (together with a further minor assumption), where $[a,b]=ab-ba$, is either associative or commutative. The present result is an anticommutative analog of this theorem.

REVIEWS

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

We invite readers to submit reviews of significant recent college-level mathematics books. We especially encourage reviews based on classroom use and comparative reviews of several related books. Reviews should ordinarily not exceed two pages (per book) typed double spaced. Manuscripts of reviews as well as books submitted for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, MN 55057.

A Short Calculus, Third Edition. By Daniel Saltz. Goodyear, Pacific Palisades, California, 1980. xii + 560 pp. $20.95.

Having used *A Short Calculus*, Second Edition, for two years in a two-semester course, I find that the third edition only enhances an already outstanding text. The positive features of the text are many. The style of writing is lucid, readable, and mathematically accurate. All of the standard topics have been covered to meet the needs of students in a liberal arts, biology, or business curriculum. The mixture of theory and application is rather evenly balanced; deep proofs are not stressed, and the calculus is developed intuitively without undue formality or excessive rigor.

The level of presentation and style of writing are accessible to a broad audience, in particular to those for whom the book is intended—freshmen and sophomores with two years of high school algebra or the equivalent.

Minor flaws have been eliminated in the revised edition of *A Short Calculus*. More business and finance problems have been added. Furthermore, references to the wide variety of examples relating to biology-medicine, business-economics, psychology, and other fields, are listed on the inside front and back covers of the text. Other improvements include more related-rate problems, and more prominence given to limits, continuity, and exponential and logarithmic functions. The new edition contains more numerical approximations of integrals, more about differential equations, and new sections on multivariable calculus.

Saltz's text compares favorably with those of his competitors. The text does not rely upon excessive graphs and charts, color illustrations, outside manuals, word or symbol lists, or gimmickry to make it more marketable, but relies on a clear presentation of the standard calculus topics. In short, this book is a truly outstanding option for instructors of a calculus course designed for students of business and the life sciences.

JOYCE LONGMAN, Villanova University

Dictionnaire des mathématiques. By A. Bouvier and M. George, under the direction of F. Le Lionnais. Presses Universitaires de France, 1979. 832 pp.

Like other specialized dictionaries, mathematical dictionaries are rare, require much work to compile, and probably earn their publishers and authors little profit. They also commonly disappoint the users. In the library of my institution, I noted only three other mathematical dictionaries of recent vintage that are comparable to the book being reviewed: *Mathematisches Wörterbuch* (*MW*), edited by J. Nass and H. L. Schmid, Akademie-Verlag GMBH, 1961; *Encyclopaedic Dictionary of Mathematics for Engineers and Applied Scientists* (*EDMEAS*), edited by I. N. Sneddon, Pergamon Press, 1976; and *Encyclopedic Dictionary of Mathematics* (*EDM*), Mathematical Society of Japan, edited by S. Iyanaga and Y. Kawada, The MIT Press, 1977. *EDMEAS* specializes in applied mathematics. *MW*, while much bigger than *Dictionnaire des mathématiques* (*Ddm*) is dated by now in some respects. *EDM* is also much bigger than *Ddm*. *EDM* organizes the material into long articles and has a very complete index. *Ddm* organizes its material into short articles with cross-referencing.

On the whole, *Ddm* would probably be a useful book for the contemporary mathematician. The definitions are as concise as possible. *Ddm*, together with the three dictionaries above, covers much of modern mathematics. A sampling of articles from *Ddm* includes: contraction mapping theorem, category, cohomology, codimension, differential manifold, Field's medalists, filter, functor, fractal objects, Hausdorff distance, Hilbert's problems, lattice, Lie group, morphism, martingale, proper value, sesquilinear, Reuleaux, and syzygies. Missing are Bourbaki, fibre bundle, Hausdorff dimension, spectrum, and spline. There is a sixteen-page list of notation and a table of random numbers. Many mathematical objects bear the names of their finders. For this reason, the dictionary contains a long list of mathematicians with their dates. The youngest noted was G. Choodnovsky, born in 1952. The oldest noted was Pythagoras of the sixth century, B.C. The quality of the articles is not uniform. For example, the cone is discussed only in R^3. The physical size of the dictionary, its printing, figures, probably its price, and the conciseness and number of definitions make the dictionary attractive.

WILLIAM A. BEYER, Los Alamos Scientific Laboratory

MISCELLANEA

51. Plato perhaps had the idea that a body cannot go from rest to a positive velocity . . . without passing through all smaller velocities.

—Galileo Galilei, Discorsi e dimostrazioni matematiche intorno à due nuove scienze, 1638 (Opere, 1898, vol. 8, p. 283).

Note: If we assume that Galilei thought of velocity as being continuous, he is stating the intermediate value property. If we take his statement at face value, he is asserting the more sophisticated proposition that a derivative, continuous or not, has the intermediate value property.

R.P.B.

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, P. Developing Mathematics in Third World Countries. Ed: M.E.A. el Tom. Math. Stud., V. 33. North-Holland, 1979, xi + 207 pp, $26.75 (P). [ISBN: 0-444-85260-3] Proceedings of the International Conference, Khartoum, March 6-9, 1978. Valuable historical document reporting on first opportunity for third world mathematicians to gather and exchange ideas on development of mathematics. Topics considered: school mathematics, university mathematics, mathematics and development, mathematics policy and international cooperation. Includes 16 invited papers, and a final report with recommendations for action. PJ

General, P. Transactions of the Moscow Mathematical Society, 1980, Issue 1. Ed: Ben Silver. AMS, 1980, v + 289 pp, $69 (P). Translation of Volume 37, 1978; eight papers on analysis and algebra. LAS

General, L. International Bibliography of Journals in Mathematical Education. Gert Schubring, Jutta Richter. Universität Bielefeld (Institut für Didaktik der Mathematik, Universitätsstrasse, 4800 Bielefeld 1), 1980, 108 pp, (P). Summary information on about 300 math education journals from around the world: name, publisher, editor, subscription address, language; and brief abstract of contents. Volume 11 of the KID (Communication-Information-Documentation) project of IDM. LAS

General, P. Mathematical Reviews Index, Volumes 57 and 58 (1979). Ed: J.L. Selfridge. AMS, 1980, $120 set (P). Part 1: Author/Key Index, 1260 pp; Part 2: Subject Index, 1015 pp. Also titled Index of Mathematical Papers, Volume 11. Contains subject and author indices to over 30,000 articles and books reviewed by Mathematical Reviews in 1979. LAS

Basic, S(13-14), L. A Guide to Metrics. Terry Richardson. Prakken Pub, 1978, vii + 199 pp, $12 (P). [ISBN: 0-911168-38-9] For self-study of metric systems by individuals or groups, high school and beyond. Emphasizes involvement and investigation; self-test exercises, with answers. Thinking metric central, not conversion. Seven basic metric units considered in 18 comprehensive chapters; including history. Eight useful appendices on work-related and consumer topics. Glossary and bibliography. Valuable resource for secondary science and mathematics teachers. PJ

Precalculus, T(13: 1). Algebra and Trigonometry for College Students. Nancy Myers. D. van Nostrand, 1980, xii + 703 pp, $16.95 (P). [ISBN: 0-442-25758-9] Standard algebra and trigonometry topics presented by formula and example, with proofs relegated to chapter ends. Review topics introduced as needed and all answers provided. Thirteen of thirty-eight units cover trigonometry, concentrating on triangle approach. Test manual available with four tests per unit. MW

Education, S(15), P. Arithmetic and Learning Disabilities: Guidelines for Identification and Remediation. Stanley W. Johnson. Allyn, 1979, xvii + 332 pp, $15.95; $8.95 (P). [ISBN: 0-205-06444-2; 0-205-06504-X] For both elementary teachers and teacher specialists to assist in identifying children with arithmetic-related learning disabilities and to aid in preparing and presenting activities which permit improved learning. In two parts: (1) theory relating learning disabilities and the arithmetic curriculum and (2) materials keyed to specific levels and content areas. PJ

Education, T*(13, 15: 1), S. Exploring Elementary Mathematics: A Small-Group Approach for Teaching. Julian Weissglass. Freeman, 1979, xviii + 289 pp, $14.95 (P). [ISBN: 0-7167-1027-7] A text or laboratory supplement for mathematics courses for elementary teachers or courses for overcoming math anxiety. Designed for small-group, laboratory approach, not lecture. Uses attribute blocks, colored rods, geoblocks, geoboard, mira, pattern blocks, etc. Includes arithmetic, number systems, geometry, probability and statistics. Excellent; deserves serious consideration. PJ

Education, T(15-16), S. Teaching Secondary Mathematics through Applications, Second Edition. Herbert Fremont. Prindle, 1979, x + 342 pp, $10. [ISBN: 0-87150-256-9] Extensive revision of first edition. Maintains original goal, teaching teachers to teach mathematics through its applications in a problem solving approach. Updated applications to "our limited resources" and "space exploration." PJ

Education, S(15), P. The Cognitive Method: A Strategy for Teaching Word Problems. Shraga Yeshurun. NCTM, 1979, ix + 50 pp, $4 (P). [ISBN: 0-87353-140-X] An unnecessarily complicated presentation of an extension of a much used strategy, "make a table," for solving traditional word problems, problems currently under attack by many writers. A controversial strategy, by the author's admission,

which accepts mechanization of problem solving once the method of solution for a specified set of verbal problems has been raised to the level of understanding. PJ

Education, S(17-18), P*. Critical Variables in Mathematics Education: Findings From a Survey of the Empirical Literature. E.G. Begle. MAA, 1979, xxvi + 165 pp, $8. [ISBN: 0-88385-430-9] A complete survey of all variables (teacher, curriculum, student, environment, instructional, testing, problem solving) studied for their effects on mathematics education. Numbers and kinds of studies during 1960-1976, reviews of literature, bibliographies, author opinions concerning directions of findings, need for further research. Emphasis on factual information at expense of expert opinion and common sense. An important beginning of a solid foundation for a theory of mathematics education. Required reading for all mathematics educators. PJ

Education, S(15), P. Informatics and Mathematics in Secondary Schools: Impacts and Relationships. Ed: D.C. Johnson, J.D. Tinsley. North-Holland, 1978, x + 158 pp, $22. [ISBN: 0-444-85160-7] Proceedings of the IFIP Working Conference, Bulgaria, September 1977. 20 papers and discussions on "informatics" (the study of computer and information processing) at the secondary school level and its impact on the mathematics curriculum, mathematics teaching and teacher training. An international look at the place and value of the computer in modern secondary education. PJ

History, S*(13-16), L. The Computer from Pascal to von Neumann. Herman H. Goldstine. Princeton U Pr, 1980, xi + 378 pp, $6.95 (P). [ISBN: 0-691-02367-0] Paperback edition of the 1972 hardcover volume (TR, June-July 1973). LAS

Foundations, P, L. Wittgenstein's Lectures. Rowman & Littlefield. Cambridge, 1930-1932. Ed: Desmond Lee. 1980, xvii + 124 pp, $15.50 [ISBN: 0-8476-6201-2]; Cambridge, 1932-1935. Ed: Alice Ambrose. 1979, xi + 225 pp, $19.50 [ISBN: 0-8476-6151-2]. A philosophical subject such as Wittgenstein's philosophy of mathematics exists only in the formulation given to it by Wittgenstein himself. (Contrast this with a mathematical subject such as set theory, which is best learned from modern sources. Only the historian of mathematics reads Cantor's original works today.) The above two books contain write-ups of Wittgenstein's lectures for 1930-1935, reconstructed from notes taken by several of his students (including the editors). For the philosopher they will provide a rich new source of evidence for understanding Wittgenstein's thought and its development. (In this period Wittgenstein first repudiated his earlier joint work with Russell.) The main theme of the Lee volume is the nature of symbols, with excursions into symbolic logic and philosophy of mathematics. The Ambrose volume contains further discussions of language, grammatical rules, and a critique of logical atomism, among other things. The absence in the Ambrose volume of an index or other topical guide to the contents is regrettable. GHM

Foundations, P. Logic Matters. P.T. Geach. U of Calif Pr, 1980, xii + 335 pp, $5.95 (P). [ISBN: 0-520-03847-9] Paperback edition. (TR hardcover edition, March 1973.) A wide-ranging collection of essays on logic and its application to fundamental philosophical questions. MB

Foundations, T(15-17: 1). Introduction to the Foundations of Mathematics, Second Edition. Raymond L. Wilder. Krieger, 1980, xvi + 327 pp, $19.50. [ISBN: 0-89874-170-X] Unaltered reprint of the second (1965) edition of this well-known text, first published in 1952. Although its approach is necessarily somewhat dated, it remains a useful introductory text, one of the few in its field. LAS

Combinatorics, S(17-18), P. Matroids and Combinatorial Geometries. T. Brylawski, D. Kelly. (Dept. of Math., U. of No. Carolina at Chapel Hill), 1980, 149 pp, $6 (P). An introductory sketch of theory of matroids and combinatorial geometries, at a relatively elementary level, occupies nearly half the volume. Applications and areas of research are the subjects of the remainder. SS

Algebra, T(18). Polynomial Identities in Ring Theory. Louis Hale Rowen. Pure and Appl. Math., V. 84. Acad Pr, 1980, xx + 365 pp, $39.50. [ISBN: 0-12-599850-3] Covers aspects of polynomial identity (PI) rings: general structure, central simple algebras. Noetherian PI-rings, extensions of PI-rings, free rings, rational identities. Contains many exercises, open problems, historical notes, large bibliography. SG

Algebra, P. Lecture Notes in Mathematics-804: First Order Algebraic Differential Equations, A Differential Algebraic Approach. Michihiko Matsuda. Springer-Verlag, 1980, vii + 111 pp, $9.80 (P). [ISBN: 0-387-09997-2] Investigates movable singularities of first order algebraic differential equations over an arbitrary differential field. Describes recent work of the author and Nishioka as well as earlier results of Kolchin and Rosenlicht. SG

Differential Equations, T(16-18: 1, 2), S*, P*, L*.** Introduction to Partial Differential Equations and Hilbert Space Methods. Karl E. Gustafson. Wiley, 1980, xv + 270 pp, $19.95. [ISBN: 0-471-04089-4] In two parts (The Usual Trinities; Fourier Series and Hilbert Space) each containing enough material for a semester course. Cleverly orchestrated first part covers operators (potential, diffusion, wave), problems (BVP, IVP, EVP), questions (existence, uniqueness, stability), boundary conditions (Dirichlet, Neumann, Robin), solution methods (separation of variables, Green's Function, variational), tools (divergence theorem, inequalities, convergence theorems), physical techniques (conservation principles, linearization assumptions, perturbation methods) and much more. Sprightly written in the "pedagogical style of...once...twice...and then, again." Text is lean but meatily foot-noted. Good indexes (author and subject); selected answers, hints and solutions.

Developed for a full-year senior-graduate course, author claims success with wide spectrum of students. Elementary systems of ordinary differential equations are a prerequisite. This exciting text is for the venturesome, but knowledgeable, instructor. JK

Differential Equations, P. Lecture Notes in Mathematics-819: Global Theory of Dynamical Systems. Ed: Z. Nitecki, C. Robinson. Springer-Verlag, 1980, ix + 499 pp, $29.50 (P). [ISBN: 0-387-10236-1] Proceedings of a conference held at Northwestern University, Evanston, Illinois, June 18-22, 1979. JAS

Functional Analysis, P. Operator Inequalities. Johann Schröder. Math. in Sci. and Eng., V. 147. Acad Pr, 1980, xvi + 367 pp, $39.50. [ISBN: 0-12-629750-9] From the Preface: "The main interest is in results that allow one to derive properties of an unknown element (vector, function,...) from operator inequalities that involve this element." The two major subjects treated are inverse-positive linear operators and monotone-nonlinear operators with results applied primarily to ordinary differential operators. Extensive notes and bibliography. SES

Functional Analysis, T(16-17: 1). Functional Analysis, A Short Course. Edward W. Packel. Krieger, 1980, xvii + 172 pp, $12.50. [ISBN: 0-89874-019-3] Unaltered reprint of the 1974 original Intext edition (TR, June-July 1974). LAS

Optimization, S(17), P, L. Combinatorial Optimization. Ed: Nicos Christofides, et al. Wiley, 1979, x + 425 pp, $47. [ISBN: 0-471-99749-8] An edited collection of papers on combinatorial optimization. Includes chapters which describe methodologies and results of general applicability to the field and chapters which discuss some of the best known problems. Based on a summer school on combinatorial optimization in SOGESTA at Urbino, Italy. CEC

Analysis, P. Lecture Notes in Mathematics-787: Potential Theory, Copenhagen 1979. Ed: C. Berg, G. Forst, B. Fuglede. Springer-Verlag, 1980, viii + 319 pp, $19.50 (P). [ISBN: 0-387-09967-0] Contains about half the papers and a list of problems from the colloquium held in Copenhagen, May 14-18, 1979. JAS

Algebraic Geometry, P. Lecture Notes in Mathematics-802: Etude Géometrique des Espaces Vectoriels II, Polyèdres et Polytopes Convexes. Jacques Bair, René Fourneau. Springer-Verlag, 1980, vii + 283 pp, $17.20 (P). [ISBN: 0-387-09993-X] Rigorous treatment of convex polyhedra and polytopes from a geometric point of view. New results on the characterization of infinite dimensional convex polyhedra, the various criteria for separation of polyhedra, and Choquet simplexes. Final chapter reviews results since 1976. Extensive bibliography. JG

Algebraic Geometry, S(18), P. Structure Sheaves over a Noncommutative Ring. Jonathan S. Golan. Pure and Appl. Math., V. 56. Dekker, 1980, xv + 170 pp, $24.75 (P). [ISBN: 0-8247-1178-5] "For the most part this work is expository...gathering together the various results scattered in the literature which 'set the stage' by introducing the noncommutative analogs of the spectrum of a ring. ...It ends with the definition of structure sheaves and ringed spaces." First of a series dealing with these concepts and their geometric meaning. Bibliography, indexes. JS

Algebraic Geometry, P. An Introduction to the Theory of Special Divisors on Algebraic Curves. Phillip A. Griffiths. CBMS Reg. Conf. in Math., No. 44. AMS, 1980, v + 25 pp, $5.60 (P). [ISBN: 0-8218-1694-2] Notes from an NSF-CBMS Conference held at the University of Georgia in 1979. Gives a partial overview of the subject (Clifford's theorem, Brill-Noether matrix, Riemann-Roch, the Jacobian) deferring more technical statements and proofs to the papers of Arbarello, Cornalba, Harris, and the author. SG

Differential Geometry, P. Manifolds of Differentiable Mappings. P.W. Michor. Shiva Pub, 1980, iv + 158 pp, (P). [ISBN: 0-906812-03-8] An exposition of the basic results which can be proved without the help of a hard implicit function theorem on nuclear function spaces. Approximately half the book is preparatory material starting with jet bundles and manifolds with corners. Short subject index, but lacks both an extensive bibliography and a much needed index of symbols. JAS

Geometry, S*(13-16), P, L*. Symmetry, An Analytical Treatment. J. Lee Kavanau. Science Software Systems, 1980, xxx + 649 pp, $24.95 (P). [ISBN: 0-937292-00-1] A labor of love. Over 600 pages, author-typed for direct photocopying. Non-traditional. Striking new results on symmetry and classification of curves. Accessible, but not without hard work, to readers with only analytical geometry and some differential calculus. Key role played by author's "circumpolar intercept transform" applied to plane curves about specific points or representative points of point arrays. Read this book for more in symmetry than meets the eye. JK

Algebraic Topology, P. C*-Algebra Extensions and K-Homology. Ronald G. Douglas. Annals of Math. Stud., No. 95. Princeton U Pr, 1980, vi + 83 pp, $12.50; $4.50 (P). The Hermann Weyl Lectures given by the author at the Institute for Advanced Study, Princeton, New Jersey, in February 1978. The algebra of extensions of the C*-algebra of compact operators by the algebra of continuous complex-valued functions on a compact metrizable space is studied as a K-homology. JAS

Topology, T?(16-17: 1, 2), S*, L*. Seifert and Threlfall: A Textbook of Topology. H. Seifert, W. Threlfall. Trans: Michael A. Goldman. Pure and Appl. Math., V. 89. Acad Pr, 1980, xvi + 437 pp,

$39.50. [ISBN: 0-12-634850-2] A translation of the 1934 German classic together with a translation of Seifert's 1933 paper "Topologie dreidimensionales gefaserter Raum." Although its age and paucity of problems make it questionable as a text, its wealth of examples, intuition, and discussion make it a boon to students who don't read German. JAS

Topology, P. Lecture Notes in Mathematics-788: Topology Symposium, Siegen 1979. Ed: U. Koschorke, W.D. Neumann. Springer-Verlag, 1980, viii + 495 pp, $29.50 (P). [ISBN: 0-387-09968-9] Proceedings of the symposium held at the Gesamthochschule of Siegen, June 14-19, 1979. Topics fall in three special areas: differential and geometric topology, equivariant topology, and homotopy theory and algebraic topology. JAS

Topology, T, S, L**.** A First Course in Algebraic Topology. Czes Kosniowski. Cambridge U Pr, 1980, viii + 269 pp, $15.95 (P); $44.50. [ISBN: 0-521-29864-4; 0-521-23195-7] The author begins his Preface with "This book provides a variety of self-contained introductory courses on algebraic topology for the average student. It has been written with a geometric flavor and is profusely illustrated (after all, topology is a branch of geometry)." Indeed the book does this very well. Point set topology is followed by the fundamental group, the Seifert-Van Kampen theorem, and applications to knot theory. A final chapter introduces singular homology. JAS

Computer Programming, T*(13), L*. Programming in BASIC for Personal Computers. David L. Heiserman. P-H, 1981, xii + 333 pp, $17.95; $7.95 (P). [ISBN: 0-13-730747-0; 0-13-730739-X] A very elementary yet detailed and really readable introduction to Basic in the context of bare bones home computer (e.g., tape cassettes rather than diskette). This really is a programming text appropriate for the serious beginner. No gimmicks, no games, but a real explanation of what's going on and why in basic Basic. Danger: The author confuses IF...THEN and ONLY IF...THEN; of course so, in a sense, does Basic. JAS

Computer Science, T*(16-17: 2), P. Computer Systems Architecture. Jean-Loup Baer. Computer Sci Pr, 1980, xiii + 626 pp, $22.95. [ISBN: 0-914894-15-3] Develops the general concepts while drawing on current architectures as examples; then discusses specific architectures. Covers the whole spectrum of architectures from micros to super computers. Historical survey. Exercises. Extensive bibliography. JL

Computer Science, T(16-17: 1), P. Principles of Firmware Engineering in Microprogram Control. Michael Andrews. Computer Sci Pr, 1980, xv + 347 pp, $21.95. [ISBN: 0-914894-63-3] On digital system design with special emphasis on micro-programmable control units. Contains detailed discussions of actual microcontrollers. For those with background in computer organization and architecture. Exercises. JL

Computer Science, S, P, L. Lecture Notes in Computer Science-89: Computer Aided Design, Modelling, Systems Engineering, CAD-Systems. Ed: J. Encarnacao. Springer-Verlag, 1980, xiv + 459 pp, $27 (P). [ISBN: 0-387-10242-6] The lectures from the advanced course on computer aided design held at the Technical University of Darmstadt from September 8-19, 1980. Much of the material is of an expository nature describing problems and approaches to solutions. This may be of value for curriculum discussion. JAS

Applications (Chemistry), P. Dynamics and Modelling of Reactive Systems. Ed: Warren E. Stewart, W. Harmon Ray, Charles C. Conley. Acad Pr, 1980, xi + 413 pp, $27.50. [ISBN: 0-12-669550-4] Proceedings of the seminar held at the University of Wisconsin, October 22-24, 1979. JAS

Applications (Engineering), T(16-17: 2), L. Introduction to Control Theory, Including Optimal Control. D.N. Burghes, A. Graham. Ellis Horwood, 1980, 400 pp, $67.95. [ISBN: 0-85312-181-8] Electrical and control engineers are the principal audience. Emphasis is on applications, with efforts to minimize mathematical complications, but use is made of differential equations, Laplace transformations, complex function theory, linear algebra, and the calculus of variations. Complete solutions to problems are provided. AWR

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; WC: William Carlson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK T. KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

Assistant Professor *Steven C. Althoen* of the University of Michigan-Flint has been promoted to Associate Professor.

Harold Boas, formerly at MIT, has been named Assistant Professor of Mathematics at Columbia University.

Barbara J. Bulmahn has been promoted from Instructor to Assistant Professor at Indiana University-Purdue University at Fort Wayne.

Associate Professor *Rebekka Struik* of the University of Colorado has been promoted to Professor.

Assistant Professor *John Cruthirds* of Pembroke State University has accepted an appointment as Assistant Professor in the Department of Mathematics and Statistics at the University of South Alabama.

Assistant Professor *Juliana Dowell* has been promoted to Associate Professor of Mathematics at Atlantic Christian College.

Associate Professor *Stanley Grossman* of the University of Montana has been promoted to Professor.

Palle Jorgensen, formerly of the Matematisk Institut Aarhus, Denmark, has been appointed Assistant Professor at Stanford University.

Associate Professor *Gary L. McGrath* of the Pittsburg State University has been promoted to Professor.

Professor *Paul R. Meyer*, Chairman of the Department of Mathematics at Herbert H. Lehman College, will be visiting at University of Padua, Italy, for the spring semester, 1981.

Eloise Taylor, Associate Professor of Mathematics at the University of District of Columbia, retired in 1980.

Stephen A. Wiitala, formerly at Nebraska Wesleyan University, has been appointed Assistant Professor at Norwich University.

Mike Zeller was appointed Assistant Professor of Mathematics and Computer Science at DePauw University.

At the University of Kentucky Associate Professors *Paul Eakin* and *Lawrence Harris* have been promoted to Professor.

At San Diego State University Associate Professors *Nicholas A. Branca* and *Douglas B. McLeod* have been promoted to Professor. *Charles R. Burton* has retired with the rank of Professor Emeritus. *Peter Salamon*, formerly of Arizona State University, has been named Visiting Assistant Professor in Mathematics.

John Carlson and *George Downing*, Associate Professors of Mathematics at Emporia State University, have been promoted to Professor.

At California State University, Chico, Professor *Harry Pollard* of Purdue University was Distinquished Visiting Professor for the fall semester, 1980. Professor *Donald Crowe* of the University of Wisconsin is Distinguished Visiting Professor for the spring semester, 1981.

At Beloit College Associate Professor *Philip Straffin* has succeeded Professor *John Finch* as Chairman of the Department of Mathematics.

At the Rose-Hulman Institute of Technology Assistant Professor *Ralph P. Grimaldi* has been promoted to Associate Professor. *Gary J. Sherman* is on sabbatical leave at Clemson University for the current academic year.

At Duke University Professor *Francis J. Murray* has retired with emeritus rank. Assistant Professor *Eric Schechter* has resigned to take a position at Vanderbilt University.

At the University of Saskatchewan *I.E. Leonard III* has been appointed Assistant Professor. He was formerly a Research Associate at Lakehead University. Associate Professor *K. Wayne Welsh* has resigned and is now at Fraser Valley College.

Louisiana State University has announced the appointment of *James K. Deveney* of Virginia Commonwealth University as Visiting Associate Professor. New Assistant Professors are *Charles N. Delzell* and *Lane Clark*. *K. Brooks Reid* and *Douglas Curtis* have been promoted from Associate Professor to Professor. *Luther I. Wade*, *Dan Scholz*, and *Richard D. Anderson* retired in May 1980 and now hold the title of Professor Emeritus.

At Darthmouth College Associate Professors *Thomas F. Bickel*, *Kenneth P. Bogart*, *James E. Baumgartner* and *Stephen J. Garland* have been promoted to Professor.

Associate Professors *Richard Carmichael* and *W. Graham May* of Wake Forest University have been promoted to Professor. *E. Lee May*, Jr., of Salisbury State College has been appointed Visiting Associate Professor.

At Northwestern University, *Sandy Zabell*, formerly of Rutgers University, has been appointed Associate Professor. *Eric Friedlander* has been promoted from Associate Professor to Professor. Professor *Ralph P. Boas* has retired with the title Henry S. Noyes Professor Emeritus.

Recent appointments at the University of Notre Dame include Associate Professor *William G. Dwyer*, formerly of Yale University, and Visiting Assistant Professors *Samuel Kleinerman* of Northwestern University and *George R. Bradley* of Notre Dame.

At Western Illinois University *Michael Ingrassia*, formerly at the University of Illinois, has been appointed Assistant Professor. *Larry Morley*, Associate Professor was named to chair the Department of Mathematics in July 1980.

At Alfred University, Chairman *Roger H. Moritz* has been promoted from Associate Professor to Professor. Associate Professor *Robert E. Ehrlich* has retired. *Robert C. Williams*, awarded a grant by N.S.F., is studying at Battelle Pacific Northwest Laboratories in Richland, Washington.

Associate Professor *Anthony Hoffman* has resigned from the faculty of S.U.C.-Genesco to take a position with the Combustion Engineering Corporation of Hartford, Connecticut.

At Randolph-Macon Woman's College Assistant Professor *Paul L. Irwin* has been promoted to Associate Professor. Assistant Professor *Beth G. Barnwell* has resigned and Professor *M. Gweneth Humphreys* has retired with the title Charles A. Dana Professor of Mathematics Emeritus.

Associate Professor *Donald F. Shriner* of Frostburg State College has been promoted to Professor.

Chicago State University: Dr. *Guang-Nay Wang*, formerly of Northwest Missouri State University, has been appointed Assistant Professor. Assistant Professor *Nancy Johnson* has been promoted to Associate Professor.

Illinois State University: Associate Professor *John Dossey* has been promoted to Professor. Assistant Professor *Stanley Seltzer* has accepted a position at Carleton College.

St. Vincent College: Bro. *David Carlson* has been appointed Lecturer, effective fall 1981. The Department of Mathematics has recently set up a five-year cooperative program in engineering with the University of Pittsburgh.

University of San Diego: Dr. *Janice B. Koop* has been appointed Assistant Professor. Assistant Professor *Lynne Small* was promoted to Associate Professor.

Associate Professor *Sidney Hsieh-Ling Kung* of Jacksonville University has been promoted to Professor.

Assistant Professor *Gary N. Baldwin* has been promoted to Associate Professor at the Lancaster Campus of Ohio State University.

Dr. *Gary Grabner* of Ohio University has been appointed Assistant Professor at the Shenango Valley Campus of the Pennsylvania State University.

Professor *Peter G. Anderson*, formerly Chairman of the Department of Computing and Decision Sciences at Seton Hall University, has been named a professor in the School of Computer Science and Technology at the Rochester Institute of Technology.

Assistant Professor *Eugene Anderson*, formerly of West Virginia Wesleyan College, has accepted a position at Indiana University at South Bend.

Erben Cook, Jr., Professor Emeritus at Central Connecticut State College, died April 23, 1980. He was a member of the Association for 19 years.

Professor *James O. Danley* of East Central Oklahoma State College died May 12, 1980 at the age of 49. He was a member of the Association for 23 years.

William R. Deane of Cincinnati, Ohio, died June 2, 1980, at the age of 62. He was a member of the Association for 8 years.

Professor *Morton J. Hillman* of Long Island University's Brooklyn Center died October 14, 1980 at the age of 62. He was a member of the Association for 31 years.

Wade H. Long of Easley, S.C., died October 4, 1980, at the age of 41. He was a member of the Association for 2 years.

Professor Emeritus *Renke G. Lubben* of San Antonio, Texas, died June 30, 1980, at the age of 82. He was a member of the Association for 56 years.

Sidney Penner of the Bronx, New York, died in May 1980 at the age of 52. He was a member of the Association for 25 years.

NORTH CENTRAL SUMMER SEMINAR

The North Central Section of MAA will hold its 3rd biennial Summer Seminar at Saint Mary's College, Winona, MN, June 22-26, 1981. The topic will be Operations Research and the program will take the form of three short courses, *Linear Programming*, *Dynamic Programming*, and *Queueing Theory and Simulation*, and a seminar in *Current Optimization Topics*. For information write to:

Louis A. Guillou
Department of Mathematics and Statistics
Saint Mary's College
Winona, MN 55987

REPORT OF THE TREASURER FOR THE YEAR 1979

Herewith is a summary of the report of the Treasurer on the operating funds of the Association. The Association also handles several grant funds as well as the funds for the High School Contests.

The full report has been approved by the Finance Committee and accepted by a vote of the Board of Governors. Any member of the Association who wishes to have a copy of the full report may obtain one by writing to the Washington Office of the Association.

ASSETS	December 31, 1979
Cash	$ -
Accounts receivable	97,639
Publication inventory	128,164
Prepaid expenses	50,753
Investments: Stocks and Bonds (at cost)	287,946
Investments: Real Estate (at cost)	237,601
Furniture & equipment (at cost)	85,905
Accumulated depreciation	(44,408)
Deferred publication costs	108,015
Total Assets	$951,615

LIABILITIES AND FUND BALANCES	
Accounts payable	$ 72,484
Accrued liabilities	15,522
Unearned dues and subscriptions	455,373
Bank note	57,000
Unearned advertising and other	7,120
Fund balances	344,116
Total Liabilities and Fund Balances	$951,615

Fund balances, December 31, 1978	$366,159
December 31, 1979	344,116
Net increase (decrease)	$(22,043)

INCOME	
Dues	$442,859
Sales of books & pamphlets	249,057
Journal subscriptions to non-members	125,655
Space rental	80,179
Advertising	57,905
Investments	29,418
Registration fees	27,379
Contributions	21,023
Miscellaneous	14,750
Total Income	$1,048,225

EXPENSES	
Salaries & related expenses	$ 373,897
Publication production costs	293,151
Travel & meetings	78,913
Awards, contributions, dues	26,738
Employment register & CML	16,196
Rent & building operations	89,384
Postage & telephone	48,799
Printing & duplicating	61,633
Supplies	17,081
Data processing	33,447
Administrative costs charged to restricted funds	(11,602)
Miscellaneous	42,631
Total Expenses	$1,070,268
Income over (under) expenses	$(22,043)

LEONARD GILLMAN, Treasurer

CALENDAR OF FUTURE MEETINGS

Sixty-first Summer Meeting, Pittsburgh, Pennsylvania, August 17–19, 1981.
Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director

ALLEGHENY MOUNTAIN, Duquesne University, Pittsburgh, Pennsylvania, May 15–16, 1981.

EASTERN PENNSYLVANIA AND DELAWARE, Saturday before Thanksgiving.

FLORIDA, Bethune Cookman College, Daytona Beach, March 13–14, 1981.

ILLINOIS, Illinois State University, Normal, May 1–2, 1981.

INDIANA, Indiana University-Purdue University, Indianapolis, April 11, 1981.

INTERMOUNTAIN, Brigham Young University, Provo, Utah, April 10–11, 1981.

IOWA, Coe College, Cedar Rapids, April 10–11, 1981 (tentative date).

KANSAS, Benedictine College, Atchison, April 11–12, 1981.

KENTUCKY, Jefferson Community College, Louisville, April 3–4, 1981.

LOUISIANA–MISSISSIPPI, Mississippi State University, Mississippi State, February 13–14, 1981.

MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, William and Mary College, Williamsburg, Virginia, April 11, 1981 (tentative date).

METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.

MICHIGAN, Oakland University, Rochester, May 1–2, 1981.

MISSOURI, Northwest Missouri State University, Maryville, April 10–11, 1981.

NEBRASKA, University of South Dakota, Vermillion, South Dakota, April 10–11, 1981.

NEW JERSEY, Seton Hall University, South Orange, spring 1981.

NORTH CENTRAL, Mankato State University, Mankato, Minnesota, May 1–2, 1981.

NORTHEASTERN, Saturday before Thanksgiving and third week in June.

NORTHERN CALIFORNIA, University of Santa Clara, March 14, 1981.

OHIO, Miami University, Oxford, April 10–11, 1981.

OKLAHOMA–ARKANSAS, Oklahoma Christian College, Oklahoma City, March 27–28, 1981.

PACIFIC NORTHWEST, second Saturday in June. Deadline for papers six weeks before meeting.

ROCKY MOUNTAIN, Colorado College, Colorado Springs, May 1–2, 1981.

SEAWAY, first Saturday in November and Saturday in late April. Deadline for papers six weeks before meeting.

SOUTHEASTERN, University of Alabama, Birmingham, April 10–11, 1981.

SOUTHERN CALIFORNIA, first or second Saturday in March.

SOUTHWESTERN, New Mexico State University, Las Cruces, April 1981.

TEXAS, San Antonio College, San Antonio, April 10–11, 1981.

WISCONSIN, University of Wisconsin, La Crosse, late March–early April 1981.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.

AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES

AMERICAN MATHEMATICAL SOCIETY, Pittsburgh, Pennsylvania, August 18–21, 1981.

AMERICAN SOCIETY FOR ENGINEERING EDUCATION

ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.

ASSOCIATION FOR SYMBOLIC LOGIC

ASSOCIATION FOR WOMEN IN MATHEMATICS

CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES

FIBONACCI ASSOCIATION

INSTITUTE OF MATHEMATICAL STATISTICS

MU ALPHA THETA

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, St. Louis, Missouri, April 22–25, 1981.

OPERATIONS RESEARCH SOCIETY OF AMERICA, Four Seasons Sheraton, Toronto, Canada, May 4–6, 1981.

PI MU EPSILON

SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

MARCH

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 3

Archimedes—Numerical Analyst
Catenary Chain Conjectures
Pólya's Counting Theorem

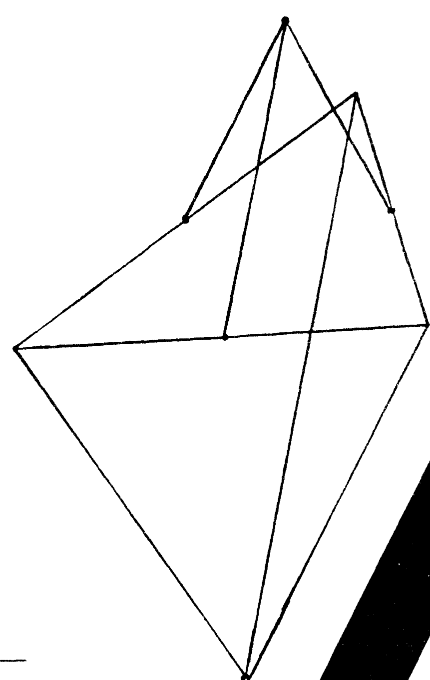
Computer Programs

Euler Lines

Strict Convexity

Binary Notation

Zeros of Polynomials



Detailed contents on cover 4

1981

THE AMERICAN
MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 3**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. (This is a change from the style previously used in this MONTHLY.) All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Research Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201; NOTES, etc.: to the corresponding Associate Editor; REPRINT PERMISSION: to R. P. BOAS (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, N.Y. 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

ARCHIMEDES THE NUMERICAL ANALYST

G. M. PHILLIPS
The Mathematical Institute, University of St. Andrews, St. Andrews, Scotland

1. Introduction. Let p_N and P_N denote half the lengths of the perimeters of the inscribed and circumscribed regular N-gons of the unit circle. Thus $p_3 = 3\sqrt{3}/2$, $P_3 = 3\sqrt{3}$, $p_4 = 2\sqrt{2}$, and $P_4 = 4$. It is geometrically obvious that the sequences $\{p_N\}$ and $\{P_N\}$ are respectively monotonic increasing and monotonic decreasing, with common limit π. This is the basis of Archimedes' method for approximating to π. (See, for example, Heath [**2**].) Using elementary geometrical reasoning, Archimedes obtained the following recurrence relation, in which the two sequences remain entwined:

$$1/P_{2N} = \tfrac{1}{2}(1/P_N + 1/p_N) \tag{1a}$$

$$p_{2N} = \sqrt{(P_{2N}p_N)}. \tag{1b}$$

We note that these involve the use of the harmonic and geometric means. Beginning with $N = 3$ and applying the recurrence formula five times, Archimedes established the inequalities

$$3\tfrac{10}{71} < p_{96} < \pi < P_{96} < 3\tfrac{1}{7}. \tag{2}$$

His skill in obtaining rational numbers $3\frac{10}{71}$ and (the very familiar) $3\frac{1}{7}$ so close to the irrational numbers p_{96} and P_{96} can be more readily appreciated if we display all four numbers to four decimal places:

$$p_{96} = 3.1410, \qquad 3\tfrac{10}{71} = 3.1408$$

$$P_{96} = 3.1427, \qquad 3\tfrac{1}{7} = 3.1429.$$

2. Stability of the Recurrence Relation. In any thorough study of a recurrence relation we need to consider the question of *numerical stability*, that is, whether rounding errors are magnified by the recurrence relation. As an example, consider the sequence $\{a_n\}$ defined by

$$a_n = \frac{2}{\pi}\int_0^{\pi} e^{\cos\theta}\cos n\theta \, d\theta. \tag{3}$$

(The a_n are the Chebyshev coefficients for e^x; see Clenshaw [**1**].) It is easily verified, on integrating (3) by parts, that this sequence satisfies the recurrence relation

$$a_{n+1} = a_{n-1} - 2na_n. \tag{4}$$

In principle, given a_0 and a_1, we may then use (4) to compute the value of any a_n. In practice, the recurrence relation (4) does not provide a satisfactory method of computing this sequence, because it is numerically unstable. To illustrate this, suppose we begin with $a_0 = 2.5321$ and $a_1 = 1.1303$, which are correct to 4 decimal places. Using (4) and rounding each a_n to 4 decimal places gives $a_2 = 0.2715$, $a_3 = 0.0443$, $a_4 = 0.0057$, $a_5 = -0.0013$, and $a_6 = 0.0187$. The true values, to 4 decimal places, are a_2 and a_3 as above and $a_4 = 0.0055$, $a_5 = 0.0005$, and $a_6 = 0.0000$. We can now see, on re-examining (4), that the error in a_{n+1} is approximately $(-2n)$ times the error in a_n, which shows why (4) is numerically unstable.

To examine the stability of (1) let us assume that, due to the effect of rounding errors, we actually compute numbers $\tilde{P}_{2N}$ and $\tilde{p}_N$ instead of P_{2N} and p_N, where

George M. Phillips did research in the theory of numbers under E. M. Wright at Aberdeen. He taught at Southampton (1963–1967) before moving to St. Andrews, where he is now Reader in Numerical Analysis. He has made research visits to the University of Texas at Austin and, on several occasions, to the University of Calgary. His main research interests are in numerical analysis and approximation theory. He is coauthor (with P. J. Taylor) of *Computers* (1969) and *Theory and Applications of Numerical Analysis* (1973).—*Editors*

$$\tilde{P}_{2N} = P_{2N}(1 + \delta), \tag{5a}$$

$$\tilde{p}_N = p_N(1 + \epsilon). \tag{5b}$$

We call δ and ϵ the relative errors in P_{2N} and p_N, respectively. To find the relative error in p_{2N}, we have

$$\tilde{p}_{2N} = \sqrt{(\tilde{P}_{2N}\tilde{p}_N)}. \tag{6}$$

Thus $\tilde{p}_{2N}$ (neglecting the rounding error incurred in evaluating the right side of (6)) is the number we would actually obtain, instead of p_{2N}. Substituting (5) into (6), we have

$$\frac{\tilde{p}_{2N} - p_{2N}}{p_{2N}} = (1 + \delta)^{1/2}(1 + \epsilon)^{1/2} - 1 \tag{7}$$

as the relative error in p_{2N}. Using binomial expansions in (7) we see that, for small values of δ and ϵ,

$$\frac{\tilde{p}_{2N} - p_{2N}}{p_{2N}} \simeq \tfrac{1}{2}(\delta + \epsilon). \tag{8}$$

An analysis of (1a) produces a result similar to (8), showing that rounding errors are not magnified by the recurrence relation, which is thus stable.

3. Rate of Convergence. We have a great advantage over Archimedes in being able to express P_N and p_N in terms of circular functions. It is easily verified that

$$p_N = N\sin(\pi/N) \tag{9}$$

and

$$P_N = N\tan(\pi/N). \tag{10}$$

From (9) and (10) we can justify that (1a) and (1b) are indeed correct and, further, from our familiarity with the Maclaurin series for $\sin\theta$ and $\tan\theta$, we can establish the rate of convergence of the sequences $\{p_N\}$ and $\{P_N\}$. Considering p_N first, we have from (9)

$$p_N = N\left[\left(\frac{\pi}{N}\right) - \frac{1}{3!}\left(\frac{\pi}{N}\right)^3 + \frac{1}{5!}\left(\frac{\pi}{N}\right)^5 - \cdots\right] \tag{11}$$

so that, for large N,

$$\pi - p_N \simeq \frac{1}{6}\pi^3 \cdot \frac{1}{N^2}. \tag{12}$$

We could give a more precise form of (12) by writing down the first two terms of the series (11) plus a remainder term. We can now see from (8) that the error in p_{2N} is approximately one-quarter of the error in p_N. More precisely, we have

$$\lim_{N\to\infty} \frac{\pi - p_{2N}}{\pi - p_N} = \frac{1}{4}. \tag{13}$$

By considering the series for $\tan(\pi/N)$, we see that the errors in the sequence $\{P_N\}$ decrease at the same rate. An inspection of the values of p_N and P_N in Table 1 shows that one might guess this result. (An explanation of the last column of this table follows later.) Given the superb numerical skills of Archimedes, one is sorely tempted to conjecture that he must have been aware of the rate of convergence of his sequences.

4. "Faster" Convergence. We have just seen that the convergence of the sequences $\{P_N\}$ and $\{p_N\}$ is very slow, and it is interesting to consider how to improve on this. First we expand (10) in a Maclaurin series to give

$$P_N = N\left[\left(\frac{\pi}{N}\right) + \frac{1}{3}\left(\frac{\pi}{N}\right)^3 + \frac{2}{15}\left(\frac{\pi}{N}\right)^5 + \cdots\right]. \tag{14}$$

TABLE 1. The first few values of p_N, P_N, and u_N.

N	p_N	P_N	u_N
3	2.598076	5.196152	3.464102
6	3.000000	3.464102	3.154701
12	3.105829	3.215390	3.142349
24	3.132629	3.159660	3.141639
48	3.139350	3.146086	3.141596
96	3.141032	3.142715	3.141593
192	3.141452	3.141873	3.141593

We may now eliminate the terms in $1/N^2$ between (11) and (14) by writing

$$u_N = \frac{1}{3}(2p_N + P_N) = \pi + \frac{1}{20}\frac{\pi^5}{N^4} + \cdots, \tag{15}$$

so that

$$u_N - \pi \simeq \frac{1}{20}\frac{\pi^5}{N^4} \tag{16}$$

and u_N converges to π faster than p_N or P_N. The first few values of u_N are given in Table 1. If we re-calculate the numbers in Table 1 to greater accuracy, we find that u_{96} gives an approximation to π which is more accurate, by a factor greater than 1000, than either of Archimedes' approximations p_{96} and P_{96}.

The technique of eliminating the term in $1/N^2$ could also have been done between p_N and p_{2N} (or, equally, between P_N and P_{2N}). Thus, similarly to (16), we can show that, say,

$$v_N - \pi = \tfrac{1}{3}(4p_{2N} - p_N) - \pi$$

also behaves like a multiple of $1/N^4$ for large N. This process is called *extrapolation to the limit*. (See, for example, Phillips and Taylor [3].) This process can be repeated; that is, we can eliminate the term in $1/N^4$ between v_N and v_{2N}. In Table 2 we show the dramatic effect of repeated extrapolation to the limit. Note that the last two numbers in the final column of Table 2 give π correct to 9 decimal places, although it is only the effect of rounding error which has prevented us from achieving agreement to twice as many places of decimals. If we re-calculate the numbers p_N in Table 2 to 20 decimal places and carry out five extrapolations (rather than three given in the table), we obtain an approximation which differs from π by less than 10^{-18}. It is remarkable that such accuracy can be extracted from Archimedes' raw material.

TABLE 2. The effect of repeated extrapolation to the limit.

N	Extrapolated Values		Repeated Extrapolation	
	p_N	v_N		
3	2.598 076 211			
6	3.000 000 000	3.133 974 596		
12	3.105 828 541	3.141 104 721	3.141 580 063	
24	3.132 628 613	3.141 561 970	3.141 592 454	3.141 592 650
48	3.139 350 203	3.141 590 733	3.141 592 651	3.141 592 654
96	3.141 031 951	3.141 592 534	3.141 592 654	3.141 592 654

5. Analysis of Convergence. In this final section we analyze the behavior of the recurrence relation (1) with arbitrary positive starting values. In divorcing (1) from its geometrical context, we shall change the notation and rewrite (1) in the form

$$1/Q_{N+1} = \tfrac{1}{2}(1/Q_N + 1/q_N) \tag{17a}$$

$$q_{N+1} = \sqrt{(Q_{N+1}q_N)}, \tag{17b}$$

beginning with arbitrary q_0, $Q_0 > 0$. We examine separately the two cases $0 < q_0 < Q_0$ and $0 < Q_0 < q_0$.

Case 1. For $0 < q_0 < Q_0$ we shall write

$$\frac{q_0}{Q_0} = \cos\theta, \qquad \alpha = \frac{q_0 Q_0}{\left(Q_0^2 - q_0^2\right)^{1/2}}, \tag{18}$$

so that

$$Q_0 = \alpha \tan\theta, \qquad q_0 = \alpha \sin\theta. \tag{19}$$

Substituting (15) into (13), we easily obtain

$$Q_1 = 2\alpha \tan \tfrac{1}{2}\theta, \qquad q_1 = 2\alpha \sin \tfrac{1}{2}\theta. \tag{20}$$

It follows that

$$Q_N = 2^N \alpha \tan(\theta/2^N), \qquad q_N = 2^N \alpha \sin(\theta/2^N), \tag{21}$$

and hence the sequences $\{Q_N\}$ and $\{q_N\}$ converge to the common limit

$$\alpha\theta = \frac{q_0 Q_0}{\left(Q_0^2 - q_0^2\right)^{1/2}} \cos^{-1}(q_0/Q_0). \tag{22}$$

The "Archimedes case" corresponds to $q_0 = 3\sqrt{3}/2$, $Q_0 = 3\sqrt{3}$.

Case 2. For $0 < Q_0 < q_0$ we write

$$\frac{q_0}{Q_0} = \cosh\theta, \qquad \alpha = \frac{q_0 Q_0}{\left(q_0^2 - Q_0^2\right)^{1/2}}, \tag{23}$$

so that

$$Q_0 = \alpha \tanh\theta, \qquad q_0 = \alpha \sinh\theta. \tag{24}$$

Substituting (24) into (17), we obtain

$$Q_1 = 2\alpha \tanh \tfrac{1}{2}\theta, \qquad q_1 = 2\alpha \sinh \tfrac{1}{2}\theta.$$

It follows that

$$Q_N = 2^N \alpha \tanh(\theta/2^N), \qquad q_N = 2^N \alpha \sinh(\theta/2^N),$$

and hence the sequences $\{Q_N\}$ and $\{q_N\}$ again converge to a common limit which, in this case, is

$$\alpha\theta = \frac{q_0 Q_0}{\left(q_0^2 - Q_0^2\right)^{1/2}} \cosh^{-1}(q_0/Q_0). \tag{25}$$

As an amusing application of this last result, let us choose

$$Q_0 = 2t, \qquad q_0 = t^2 + 1$$

for any positive $t \neq 1$. Then from (25) the sequences $\{Q_N\}$ and $\{q_N\}$ have common limit

$$\frac{2t(t^2+1)}{(t^2-1)} \log t.$$

This gives a simple method for evaluating $\log t$ and repeated extrapolation may be used to accelerate convergence. However, this is not proposed as a practical algorithm for computing $\log t$.

Acknowledgments. I am indebted to my colleagues J. M. Howie and J. J. O'Connor for interesting discussions

on this topic. I also wish to thank John W. Wrench, Jr., for his valuable comments and for the trouble he took to re-work independently the calculations embodied in Tables 1 and 2.

References

1. C. W. Clenshaw, Chebyshev Series for Mathematical Functions, Mathematical Tables, vol. 5, National Physical Laboratory, H.M.S.O., London, 1962.
2. T. L. Heath, The Works of Archimedes, Cambridge University Press, 1897.
3. G. M. Phillips and P. J. Taylor, Theory and Applications of Numerical Analysis, Academic Press, 1973.

A BRIEF HISTORY AND SURVEY OF THE CATENARY CHAIN CONJECTURES

L. J. RATLIFF, JR.
Department of Mathematics, University of California, Riverside, CA 92521

1. Introduction and Some Terminology. There is a collection of problems in commutative algebra known as the catenary chain conjectures. These conjectures, some of which have their origins in W. Krull's foundational work in 1937, are concerned with the extent to which certain useful properties hold in the integral closure of a Noetherian domain. The purpose of this article is to tell what the most important of these conjectures say, where they came from, and what their current status is.

A very brief summary of these conjectures is that they are concerned with maximal chains of prime ideals in integral extension domains of Noetherian domains. This summary will be made considerably more specific in Sections 2–7, but since some of the terminology in the preceding sentence may not be familiar to the reader, the remainder of this section will be devoted to explaining some of the relevant definitions and giving some examples to illustrate them. (Other definitions will be given when they are needed in later sections of the paper.)

The conjectures and related results are a small but well-defined and important area in the study of *Noetherian rings*—those rings R which are commutative, have an identity $1 \neq 0$, and for which every ideal is finitely generated or, equivalently, that satisfy the ascending chain condition (that is, every strictly ascending chain of ideals of R is finite). These rings are named after Emmy Noether, who, in 1921 in a very important paper **[27]**, was the first to recognize their importance. They have been extensively studied ever since, and many very important and interesting theorems concerning them have been discovered. They are now clearly one of the basic structures in all of mathematics.

Actually, most of the conjectures and problems in this area can be reduced to local considerations; that is, it is sufficient to restrict attention to Noetherian rings with a unique maximal ideal. Such rings are called *local rings*, and they arise naturally in algebraic geometry (in studying the geometry on an algebraic variety in the neighborhood of a point) and in algebraic number theory (in solving Diophantine problems). (The reduction from the global conjectures [for Noetherian rings in general] to their local versions is readily accomplished by localizing at maximal ideals of R. The method of reducing global problems to local ones is standard in commutative algebra and need not be considered here.) Local rings are topological rings, the topology being given by using the set of powers of the maximal ideal M as the

The author received his Ph.D. at the State University of Iowa under H. T. Muhly in 1961. He was a lecturer at Indiana University from 1961 to 1963 and has been at the University of California, Riverside, since then. He has written extensively on commutative algebra and has a monograph, *Chain Conjectures in Ring Theory*, in the Springer Lecture Notes in Mathematics. His main interests are in commutative algebra, Noetherian lattices, algebraic geometry, number theory, and teaching.—*Editors*

fundamental system of neighborhoods of zero. (Thus the cosets $x+M^n$ with $n\geqslant 0$ [where $M^0=R$] are the basic open neighborhoods of $x\in R$.) A local ring R is also a metric space with metric given by $d(x,y)=2^{-n}$, where $x,y\in R$ and $x-y\in M^n-M^{n+1}$; the fact that $d(x,y)$ is a metric is based on Krull's Intersection Theorem, which states $\bigcap_{n>0}M^n=(0)$. An important and very useful property of a local ring is that its topological completion is again a local ring, called the *completion* of R and denoted by R^*, and R is said to be *complete* if $R=R^*$. Complete local rings have several nice properties that do not hold for local rings in general and, because of I. S. Cohen's structure theorems **[4]**, they are considerably better understood.

Prime ideals play a particularly important role in Noetherian rings; so it may be somewhat of a surprise (because of the extensive research on Noetherian rings) to find that there are many important open problems concerning *Spec* R, the set of all prime ideals of R. (Recall that an ideal P of R is *prime* if R/P is an integral domain or, equivalently, if $a,b\in R$ and $ab\in P$ imply either $a\in P$ or $b\in P$.) Spec R is a poset (partially ordered set) under set inclusion in R, and it is also a topological space (under the spectral [or Zariski] topology). In a beautiful paper in 1969 **[12]**, M. Hochster solved a long-standing open problem by characterizing Spec R (considered as a topological space) for an arbitrary commutative ring R, but Spec R, for R Noetherian, remains (embarrassingly) something of a mystery. Here, we shall only be interested in its poset properties; even so, much of the mystery remains. (Some of the mystery is being cleared away, however. For example, on pages 6–7 of **[15]**, I. Kaplansky considered what properties a poset must have in order to be Spec R [as a poset] for a Noetherian ring R, and he listed four such conditions. Then on page 169 he noted that the list had been considerably enlarged in the period 1970–1974 [between the first and revised editions of his book] and that the enlarged list was growing rapidly. Also, a number of recent papers have added to the knowledge in this area.)

The remaining definitions that we need are all concerned with chains of prime ideals. Specifically, if $P\subset Q$ in Spec R, then a chain

$$P=P_0\subset P_1\subset\cdots\subset P_n=Q \tag{1.1}$$

in Spec R is a *saturated chain of prime ideals between P and Q* if for each $i=0,1,\ldots,n-1$ there are no $p\in$Spec R such that $P_i\subset p\subset P_{i+1}$. ($I\subset J$ means I is a proper subset of the set J.) The *length* of the chain (1.1) is n (so it is the number of joins in the chain that are counted to give its length). A *maximal chain of prime ideals in R* is simply a saturated chain of prime ideals between a minimal prime ideal and a maximal (prime) ideal. If $P\in$Spec R, then the *height* of P is the supremum of the lengths of chains of prime ideals descending from P, the *depth* of P is defined analogously by using chains ascending from P, and the *altitude* of R is the supremum of the heights of the maximal ideals in R or, equivalently, the supremum of the depths of the minimal prime ideals in R. For commutative rings in general, height, depth, and altitude may not be finite. However, height P is always finite when R is Noetherian, and, in fact, height P is at most equal to the minimal number of generators of P, by the Generalized Principal Ideal Theorem **[51**, Theorem 30, p. 240]. (In passing, it should be noted that the Principal Ideal Theorem and its generalization were proved by W. Krull in 1928 in **[16]**. Commenting on the importance of these theorems, on pages 104–105 of **[29]**, D. G. Northcott said they mark a turning point in the development of the theory of Noetherian rings, and on page 104 of **[15]** Kaplansky said the Principal Ideal Theorem is probably the most important single theorem in the theory of Noetherian rings.) Moreover, altitude R and depth P are also always finite when R is a local ring, and altitude $R=$height $M\geqslant$ depth P. (In the literature, height P is often called rank P, dimension P and co-rank P are frequently used in place of depth P, and dimension R and Krull dimension R are often used instead of altitude R.)

Finally, R is said to be *catenary* if for each pair $P\subset Q$ in Spec R, all saturated chains of prime ideals between P and Q have the same length. Of course the catenary condition plays a fundamental role in everything that follows.

To give some examples of the preceding definitions, we restrict attention to the poly-

nomial ring $R=F[X,Y]$ where F is a field. (The reader can readily generalize to the case $F[X_1,\ldots,X_n]$.) Then R is Noetherian, by the Hilbert Basis Theorem [**51**, Theorem 1, p. 201], and R is a UFD (unique factorization domain); so each irreducible polynomial f generates a principal prime ideal (f), and minimal nonzero prime ideals are principal ideals. Also, $(0)\subset(X)\subset(X,Y)$ is a maximal chain of prime ideals of length two in R, height$(0)=0=$ depth(X,Y), height$(X)=1=$depth(X), and height$(X,Y)=2=$depth(0), so altitude $R=2$. Further, R is catenary, as are all finitely generated rings over a field (see (2.1)); an example of a noncatenary ring is given at the end of Section 4. Finally, $L=R_{(X,Y)}$ (the set of all quotients f/g with $f,g\in R$ and $g(0,0)\neq 0$) is a local domain whose completion is the power series ring $L^*=F[[X,Y]]$. (In this case, L^* is again an integral domain, but in general this does not hold. In fact, the minimal prime ideals in the completion of a local domain need not have the same depth and the zero ideal can have imbedded associated primes, and it is because of this that there exist open problems concerning maximal chains of prime ideals; see the comments following (2.3).)

2. Three Important Results. Further definitions will be needed below, but at this point we can at least mention three of the most important results in this area, two of which are positive and the third negative.

(2.1) *A finitely generated ring A over a field F is catenary.*

(2.2) *A complete local ring is catenary.*

(2.3) *There exist noncatenary local domains.*

(2.1) is a classical result (see below), (2.2) was proved by Cohen in 1946, in [**4**, Theorem 19], and (2.3) was proved by M. Nagata in 1956, in [**24**, Section 3]. Nagata's result answered the following question, which had been open for some time and had come to be known as the *chain problem of prime ideals*: Is every local domain catenary? (It seemed quite likely that the answer should be yes, since a local domain is a dense subspace of its completion, which *is* catenary, by (2.2).) Actually, Nagata showed, in [**24**, Theorem 1], that the answer is yes for *quasi-unmixed* local domains (local domains such that all minimal prime ideals in their completion have the same depth; for example, all local domains of algebraic number theory and classical algebraic geometry, regular local rings, and homomorphic images of Macaulay rings), but he then went on to give an explicit construction of a family of local domains which are not catenary.

Because the answer to the chain problem of prime ideals is no (that is, since there exist local domains which are not quasi-unmixed), a number of related problems concerning chains of prime ideals in integral extension domains of Noetherian domains have subsequently appeared in the literature. Some of these have recently been settled, and certain others are now called the catenary chain conjectures. Some comments on both types will be made a little later. First, however, we pause to look somewhat more closely at (2.1)–(2.3).

The foundational work in this area is Krull's 1937 paper [**18**]. However, the origins of this paper are clearly seen in Section 17 of his 1935 book [**17**], and they can even be traced to E. Noether's 1923 result that the depth of a prime ideal in a polynomial ring is equal to the geometric dimension of the associated variety [**28**, Satz 7] (and, although it is clearly Krull's work that lays the foundation and much of the superstructure in this area, it would be difficult to overestimate the fundamental importance and pervasive influence of E. Noether's results and methods in all its developments). In this 1937 paper, Krull proved the Dimensionssatz which shows, among other things, that (2.1) holds. In Footnote 8 in [**18**], he stated he knew only two proofs of this theorem; one is based on B. L. van der Waerden's 1933 multiplicity theory of finite systems of algebraic equations [**48**, p. 120], and the other is the ideal-theoretic proof in his own book [**17**]. So, although the version of this theorem in [**18**] is a considerable generalization of the version in [**17**], the proof remained essentially the same. Since (2.1) is such a fundamental result in this area, a sketch of Krull's proof of it will now be given. (For the proof and the

remainder of this paper we need to consider integral extension rings. By definition, a ring A is an *integral extension ring* of a subring B [or is *integrally dependent* on B] in case every element in A is the root of a monic polynomial over B. The *integral closure* of B is the largest integral extension ring of B contained in its total quotient ring, and B is *integrally closed* if it is equal to its integral closure.)

Since every ring of altitude zero or one is catenary, it may be assumed that $n =$ altitude $A > 1$. Also, it follows from the definition of catenary that it may be assumed that A is an integral domain. Let p be a height one prime ideal in A. If A is the polynomial ring $F[X_1, \ldots, X_n]$, then A is a UFD. Therefore p is principal; so it follows that depth $p = n - 1$. If A is not a polynomial ring, then A is integrally dependent on a polynomial subring B, by the Noether Normalization Theorem [**52**, Theorem 25, p. 200], and height $p \cap B = 1$, by the Going Down Theorem [**51**, Theorem 6, p. 262]. Also, by integral dependence and since B is a polynomial ring over F, depth $p =$ depth $p \cap B =$ altitude $B - 1 =$ altitude $A - 1 = n - 1$. From this, if $P \subset Q$ are prime ideals in A such that height $Q/P = 1$, then depth $Q =$ depth $Q/P =$ altitude $A/P - 1$, since A/P is finitely generated over F and Q/P is a height one prime ideal in A/P. Therefore, depth $P =$ altitude $A/P =$ depth $Q + 1$; so it readily follows that every maximal chain of prime ideals in A has length n, and so A is catenary.

(The reader interested in a more detailed proof is referred to [**10**, (31.16)], [**17**, Section 17], [**20**, Section 14], [**26**, Section 14], or [**52**, Chapter 7, Section 7].)

Cohen's proof that a complete local ring R is catenary [**4**, Theorem 19] follows the same broad outline. Specifically, he first proved an analog, for power series rings, of the Noether Normalization Theorem (which he noted, in [**4**, Footnote 19], extends earlier results of W. Rückert [**46**, p. 266] and Krull [**18**, Theorems 10 and 12] to finite coefficient fields), and he then proved (in Theorem 18) that the subring obtained is a UFD. Using these two results, the proof that R is catenary is very similar to the proof above that A is catenary.

Nagata's ingenious construction of noncatenary local domains, in [**24**], is a little too lengthy to consider in detail here. (The interested reader will find the examples reproduced in [**26**, Example 2, pp. 203–205], and similar examples are given in [**20**, (14.E)] and in [**52**, Example, pp. 327–329].) Suffice it to say here that for each pair of integers $r \geqslant 1$ and $m \geqslant 0$ he constructed a local domain R of altitude $r+m+1$ such that the integral closure of R in its quotient field is catenary and has exactly two maximal ideals, one of height $m+1$ and the other of height $r+m+1$. For $m=0$, R is catenary, but for $m>0$, R is not catenary and the maximal chains of prime ideals in R have length either $m+1$ or $r+m+1$. These examples have proved to be very useful when one considers other problems concerning saturated chains of prime ideals, and recently several interesting variations of them have appeared in [**1**], [**3**], [**6**], [**9**], and [**11**]. In this regard, the result in R. Heitmann's paper [**11**] is particularly noteworthy. It shows that every finite partially ordered set S is isomorphic to a saturated subset of Spec R for some Noetherian ring R (a local domain, if there exist unique maximal and minimal elements in S); so it follows that there is no finite bound on the noncatenarity of Noetherian domains. Very recently this result was strengthened and its proof was simplified in [**7**].

3. An Important New Result. Until very recently, all known examples of noncatenary local domains were closely related to Nagata's original examples, in that they all had the same two basic features; their integral closures were catenary and had at least two maximal ideals of different heights. In fact, probably the outstanding open problem in this area was whether or not the integral closure must be catenary. This problem, known as the Chain Conjecture (CC), was essentially asked by Nagata in 1956, in [**24**, Problem 3, p. 62], and it was solved late in 1979 by T. Ogoma, who showed, in [**31**], that *there exist integrally closed local domains which are not catenary*. (Actually, he proved a considerably stronger result: there exists an integrally closed pseudo-geometric Henselian local domain of altitude three which is not catenary. This gives a negative answer to the question I asked in 1977 in [**40**, p. 322], and this is the case of the

conjecture that it seemed must certainly hold, since both the pseudo-geometric and Henselian hypotheses are strong conditions to impose on a ring and since it is well known that an integrally closed local domain of altitude at most two *is* catenary, since it is a Macaulay ring.) Ogoma showed this by giving a quite explicit (and very intricate) construction of such a ring. The construction involves a complicated numbering argument as well as some new ideas developed in the paper. (Heitmann has written a very readable version of this example that uses only standard ideas in commutative algebra and, in particular, that avoids the difficult numbering argument.[1]) Ogoma's (and Heitmann's) construction is similar to the constructions of Nagata and Heitmann, mentioned in Section 2, that involved adjoining a power series to a local domain in such a way that the new ring is still a local domain. Thus this method has now been used three different times to obtain a significant result concerning chains of prime ideals; so it appears to be especially important in this area of commutative algebra.

In closing this section, it should be mentioned that there are some very good new ideas developed in Ogoma's paper, and his main result is every bit as significant and important as the three results mentioned in Section 2. It is not as yet known if the new ideas he develops in this paper will be useful in solving some of the other catenary chain conjectures.

4. Four Chain Conjectures. As just noted, the outstanding open problem in this area has just been solved, but many related problems still remain open. So in this section, four of the most important chain conjectures will be briefly discussed. (For these conjectures, an integral domain A is a GB-*domain* if saturated chains of prime ideals in integral extension domains contract in A to saturated chains of prime ideals, and A is *normal* if it is equal to its integral closure. Also, A satisfies the f.c.c. (*first chain condition*) (respectively, the s.c.c. (*second chain condition*)) if it is catenary and all maximal chains of prime ideals in A (respectively, in all integral extension domains of A) have the same length. Concerning these last two conditions, note that a local domain is catenary if and only if it satisfies the f.c.c., and Nagata's example for the case $m = 0$ and $r \geqslant 1$ shows that there exist local domains that satisfy the f.c.c. but not the s.c.c.)

AVOIDANCE CONJECTURE (AC). *If $P \subset Q \subset N$ is a saturated chain of prime ideals in a Noetherian ring A and if $N_1, \ldots, N_h$ in* Spec A *are such that $N \not\subseteq \cup N_i$, then there exists $q \in$* Spec A *such that $P \subset q \subset N$ is saturated and $q \not\subseteq \cup N_i$.*

GB CONJECTURE (GBC). *The integral closure of a Noetherian domain is a* GB-*domain.*[2]

DEPTH CONJECTURE (DC). *Every prime ideal P of height > 1 in a local domain contains a prime ideal p such that* depth $p =$ depth $P + 1$.

NORMAL CHAIN CONJECTURE (NCC). *If the integral closure of a local domain R satisfies the* f.c.c., *then R satisfies the* s.c.c.

Concerning these conjectures, note that only the AC (respectively, the DC) is specifically concerned with the prime spectrum of an arbitrary Noetherian (respectively, local) domain. This is because of Nagata's examples: that is, these examples show that some condition (such as integrally closed) in the other two conjectures is necessary. Also, it should be noted that the conjectures are related. Specifically, it is shown in [**38**, Chapter 3] that AC⇒NCC, GBC⇒NCC, and DC⇒NCC. No other implications among these specific conjectures are known. (However, it was also shown in [**38**] that the CC implies each of the others; but because of Ogoma's result, this is no longer of much interest.)

Before further discussing these conjectures, perhaps some indication should be given as to why they (and the related problems in this area) are of some interest and importance. One reason is that they are concerned with fundamental structural properties of local rings and there are quite a few open problems, some of quite long standing. Another is that many of the chain

conditions (catenary, f.c.c., s.c.c., quasi-unmixed, altitude formula [see (6.1.1)]) are inherited by all finitely generated extension domains of a Noetherian domain, and others descend to certain Noetherian sub-domains (see, for example, [**32**, Section 3]); so knowing they hold for one ring implies they hold for many related rings. But probably the main reason is that, since the various chain conditions are quite useful and nice properties for a ring to have and so are frequently hypothesized in research problems in algebraic geometry and commutative algebra, it would be very nice to know exactly the extent to which these conjectures hold.

We next note that the position of these conjectures between certain known results is quite closely fixed. Specifically, and as already noted, the CC (which is false) implies each of the others. And, on the other hand, it was proved in 1971, in [**34**, Corollary 3.13], that the following statement (weaker than the NCC) is true: if R is a catenary local domain, then R_p satisfies the s.c.c. for all nonmaximal $P \in \operatorname{Spec} R$. And in 1978, in [**38**, (14.1)], it was shown that the NCC holds for Henselian local domains and for local domains of the form $A[X]_{(P,X)}$.

The NCC has been open for quite some time. It arose in 1956 from Nagata's incomplete proof of Proposition 1a in [**24**]. The others (including those not considered here) were first specifically stated within the past nine years. However, the general case (for arbitrary integrally closed integral domains) of the GBC was considered by Krull in 1937, in [**18**, p. 755]. In 1972, in [**14**], Kaplansky showed that this version of the conjecture is false.

The AC and DC arose from considering certain known results. Namely, the AC arose from adding an additional condition (saturation) to the following well-known and useful result [**51**, Lemma, p. 240]: if $P \subset Q \subset N$ is a chain of prime ideals in a Noetherian ring A and $N_1, \ldots, N_h$ are prime ideals in A such that $N \nsubseteq \cup N_i$, then there exists a prime ideal q in A such that $P \subset q \subset N$ and $q \nsubseteq \cup N_i$. And the DC arose from trying to "invert" the following known result (proved by S. McAdam in 1974 in [**21**, Theorem 1]): if p is a prime ideal in a Noetherian ring R, then all but finitely many elements in $\{P;\ P \in \operatorname{Spec} R \text{ and } p \subset P \text{ is saturated}\}$ satisfy height $P =$ height $p + 1$. (It is shown in [**38**, (B.5.8)] that a direct inversion of this result does not hold; that is, there exists a local domain R with a prime ideal P such that infinitely many elements in $\{p;\ p \in \operatorname{Spec} R \text{ and } p \subset P \text{ is saturated}\}$ satisfy depth $p >$ depth $P + 1$.)

In [**38**], many equivalences of these conjectures are proved, and several of these sound so reasonable that it seems they certainly should be true. (However, I made a similar statement above concerning the chain problem of prime ideals and a special case of the CC.) In any case, with the extensive and deep research in Noetherian rings over the past 58 years, it seems incongruous that we do not know the validity of, for example, the DC; and an even stronger comment would seem appropriate concerning the altitude three cases of (5.1) and (5.2).

In ending the discussion of these specific conjectures, it should be noted that they all hold for complete local domains, and for all *quasi-local* domains (domains with a unique maximal ideal) of altitude at most two whose integral closures have all their maximal ideals of the same height. However, they do not hold for all quasi-local domains. For example, let S be the partially ordered set $\{a, b, c, d, e, f\}$ with $a \subset b \subset c \subset e \subset f$ and $a \subset b \subset d \subset f$. Then, by [**19**, (2.9)], there exists a quasi-local domain R such that $\operatorname{Spec} R$ is isomorphic to S, and it is readily seen that the AC and the DC fail to hold for R. (By using Heitmann's result [**11**], there exists a local domain L of altitude four whose prime spectrum has a saturated subset isomorphic to S. However, given prime ideals $P \subset Q$ in any Noetherian ring R such that height $Q/P > 1$, there exist infinitely many prime ideals in R between P and Q. Thus it cannot be immediately claimed that our local domain L does not satisfy the AC or the DC. This ready availability of prime ideals between such P and Q accounts for some of the difficulties in solving problems in this area.)

5. Two Open Problems. Besides the preceding conjectures, there are many related open problems, of which only two of the easiest to state will be mentioned here.

(5.1) *Does every local UFD satisfy the* s.c.c.?

(5.2) *Is every pseudo-geometric Henselian local domain a* GB-*domain*?[3]

The answer to each question is yes for local domains of altitude less than three, but the answer is unknown otherwise. However, concerning (5.2), it was recently shown in [**45**] that every integrally closed Henselian quasi-local domain of altitude three is a GB-domain.

6. Some Recently Solved Problems. It has already been noted that some of the open problems in this area have recently been solved. Some of these results have been mentioned above, and several others will now be noted.

(6.1) Over the years, a number of conditions on chains of prime ideals in a local domain R have been considered in the literature. In 1969, in [**32**, Theorems 3.1, 3.6, and 3.11], I showed that the following conceptually quite different ones are, in fact, equivalent: R is quasi-unmixed; R satisfies the s.c.c.; $R[X]$ is catenary; $R[X_1,\ldots,X_n]$ is catenary for all $n \geqslant 0$; every simple integral extension domain $R[b]$ of R satisfies the f.c.c.; and , R satisfies the *altitude formula* (that is, each prime ideal P in each finitely generated integral domain A over R satisfies

$$\text{altitude } A_p + \operatorname{trd}(A/P)/(R/p) = \text{altitude } R_p + \operatorname{trd}(A/R), \tag{6.1.1}$$

where $p = P \cap R$ and trd denotes transcendence degree). (In 1970, H. Seydi re-proved this and gave some related results in [**47**], and other related results were given at that time in [**33**].) Three years later, in [**35**, Theorem 3.3], I showed that these are also equivalent to (6.1.1) holding when P is restricted to lying over the maximal ideal in R. (This showed that two different definitions of the altitude formula in the literature are equivalent for Noetherian domains.) Nagata had previously considered some of the implications between the first six conditions in 1956 and 1959, in [**24**] and [**25**]; concerning this, see [**38**, (2.5)].

(6.2) In 1956, in [**24**, Problem 1, p. 62], Nagata asked if the zero ideal in the completion of a local domain R can have imbedded associated primes. Three years later, in [**25**, Section 4], he asked if $R^{(w)} = \cap\{R_p;\ p \in \operatorname{Spec} R \text{ and } p \neq M\}$ must be a finite R-algebra when the integral closure of R is quasi-local. He then commented that an affirmative answer would show that the answer to his first question is no and that the CC holds, but a negative answer would almost certainly show that the answer to the first question is yes. In 1970, in [**8**, Proposition 3.3], D. Ferrand and M. Raynaud gave an example that showed that the answer to the second question is no and the answer to the first one is yes.

(6.3) In [**24**, Problem 1, p. 62], Nagata also asked if a quasi-unmixed local domain R must be *unmixed* (that is, must *all* associated primes of zero in R^* have the same depth?). About 1970, Kaplansky asked the stronger question: is every quasi-unmixed local domain a homomorphic image of a Macaulay ring? (It is stronger by [**26**, (34.9)].) The just mentioned example in [**8**], together with [**32**, Proposition 3.5], shows that the answer to each question is no. (However, it is still unknown if a factor domain of an unmixed local domain is necessarily unmixed. This also was asked in [**24**, Problem 2, p. 62]. It is shown in [**26**, (34.5)] that R/p is always quasi-unmixed; and in 1974, in [**2**, (5.9)], M. Brodmann showed that there are at most finitely many height one p such that R/p is not unmixed, but no further progress has been made since 1974.) Concerning Kaplansky's question, it turns out that there are actually many close connections between Macaulay domains and Noetherian domains that satisfy the altitude formula. For example, in [**44**] it is shown that many known results concerning an R-sequence in a Macaulay local domain R have very close analogs that hold for the integral closure of an ideal of the *principal class* (an ideal of height h that can be generated by h elements) in a local domain that satisfies the altitude formula.

(6.4) In 1956, in [**50**], M. Yoshida asked if a local domain R that satisfies the following condition is catenary: height P + depth P = altitude R for all $P \in \operatorname{Spec} R$. (This question was clearly suggested by Cohen's 1954 paper [**5**].) An affirmative answer was given in 1972, in [**35**,

Theorem 2.2]. And in 1974, McAdam and I showed, in [**22**, Proposition 7], that this continues to hold for all local rings.

(6.5) As already mentioned, in 1974, in [**21**, Theorem 1], McAdam proved that only finitely many prime ideals immediately above a given prime ideal p in a Noetherian domain R can have height $>$ height $p+1$. In 1975, in [**13**], he and E. Houston considered several questions related to this. Therein they defined $p \subset P$ (in Spec R) to be *k-abnormal* if height $P/p = 1$ and height $P =$ height $p +$ height $P/p + k$, and then they showed that for a given p, the set $\{k$; there exists $P \in$ Spec R such that $p \subset P$ is k-abnormal$\}$ is finite and the set $\{P \in$ Spec R; $p \subset P$ is k-abnormal for some $k \geqslant 1\}$ has only finitely many minimal elements. (They left open the question if the set $\{k$; there exist $p \subset P$ in Spec R such that $p \subset P$ is k-abnormal$\}$ can be infinite. This, among other things, was recently answered affirmatively in [**1**] and [**6**].) These results add to the knowledge of the possible pathological properties of Spec R.

(6.6) At the end of two papers on GB-rings, [**37**] and [**39**], I asked several questions about them. Most of these were recently answered in [**3**, (17) and (20)]. Specifically, it was shown in [**3**] that: there exist GB-local domains which are not catenary; there exist local domains $R \subset S$ such that S is integral over R, S is a GB-ring, and R is not; and, given $n > 0$, there exists a local domain R such that $R[X_1, \ldots, X_{n-1}]$ is a GB-ring and $R[X_1, \ldots, X_n]$ is not. (These results greatly add to the knowledge of GB-rings and their relationship to catenary rings.) Related to these results and to the GBC, and as noted in Section 5, it is now known that every integrally closed Henselian quasi-local domain of altitude at most three is a GB-ring.

7. Some Properties and Characterizations. Besides the results mentioned so far, there are many characterizations of a Noetherian domain that satisfies one or another of the various chain conditions for prime ideals (Appendix A in [**38**] has a fairly complete list), and such integral domains also have some additional interesting properties. Since several of these are somewhat unexpected and touch on other areas of current research interest, three quite recent such results concerning the altitude formula, (6.1.1), will be briefly described in this section.

(7.1) In [**36**], it was shown that, if a Noetherian domain A satisfies the altitude formula, then for each ideal I of the principal class in A and for each positive integer n, the *integral closure in A of I^n*, denoted by $(I^n)_a$, has no imbedded associated primes; and, in fact, this property characterizes such rings. Closely related to this, it was shown in [**42**, Theorem 1] that A satisfies the altitude formula if and only if the following condition holds: if I is an ideal in A and $P \in$ Spec A is an associated prime of $(I^n)_a$ for all large n, then height $P = l(IA_P)$, the *analytic spread* of IA_P (see [**30**]).

(7.2) If A is an integral domain and $P \subset Q$ in Spec A, then it is said that $P \subset Q$ satisfy *going down* in case, for all integral extension domains B of A and for all $Q' \in$ Spec B such that $Q' \cap A = Q$, there exists $P' \in$ Spec B such that $P' \subset Q'$ and $P' \cap A = P$. There are many research papers in which this condition has been studied. In a very recent one, [**23**], McAdam showed the following two results: if $P \subset Q$ in Spec A do not satisfy going down, then $(P^n)_a$ has an imbedded associated prime for all large n; and, if A is a local domain that satisfies the altitude formula and, if $P \in$ Spec A is such that $l(P) =$ height P, then $P \subset Q$ satisfy going down for all $Q \in$ Spec A that contain P.

(7.3) In (6.1) it was noted that the associated prime ideals of the zero ideal in the completion R^* of a local domain R are closely related to whether or not R satisfies the altitude formula. Related to this, it was shown in [**41**, (9)] that there exists $n > 1$ such that the maximal ideal M in R is an associated prime of all nonzero ideals $I \subseteq M^n$ if and only if there exists a depth one associated prime of zero in R^*. And, in [**43**, Theorem 1], it was shown that there exists $m > 1$ such that M is an associated prime of the integral closure I_a of all nonzero ideals $I \subseteq M^m$ if and

only if there exists a depth one *minimal* prime ideal in R^*. Using these results, the class of Noetherian domains A with the following property was essentially characterized in [**43**]: for each ideal I in A, the set of associated primes of all large powers of I is equal to the set of associated primes of all the ideals $(I^n)_a$ for n large. The altitude formula plays an important role in this characterization.

8. Final Comments. Quite a few conjectures, solved problems and recently established results concerning the various catenary chain conditions have been mentioned in this paper. Many related results could be added in each of these categories, additional related topics could be included, and relationships between this area and various other areas of study in commutative algebra (such as valuation rings, analytically independent elements, big Cohen-Macaulay modules, quadratic and monadic transformation rings, Rees rings, associated graded rings, homological methods, and transcendental extension rings) could be discussed. But this paper is already sufficiently long; so it will be closed by simply noting that the reader has seen a fair variety of the concepts and a number of the important results and problems in this area, and it is hoped enough has been done to show some of its richness and beauty.

The research for this paper was supported in part by NSF Grant No. MCS 77-00951-A02.

Notes Added in Proof

1. A noncatenary normal local domain, Rocky Mountain J. Math. (to appear).
2. Using Ogoma's example, this past summer I showed in "Four Notes on GB-rings" (22-page preprint) that this conjecture does not hold.
3. See note 2.

References

1. J. T. Arnold and M. B. Boisen, Jr., Abnormalities in Noetherian rings, Proc. Amer. Math. Soc., 73 (1979) 1–6.

2. M. Brodmann, Über die minimale Dimension der assoziierten Primideal der Komplettion eines lokalen Integritaetsbereiches, Ph.D. dissertation, Univ. of Basel, Switzerland, 1974.

3. ______, A particular class of regular domains, J. Algebra, 54 (1978) 366–373.

4. I. S. Cohen, On the structure and ideal theory of complete local rings, Trans. Amer. Math. Soc. 59 (1946) 54–106.

5. ______, Lengths of prime ideal chains, Amer. J. Math., 76 (1954) 654–668.

6. A. M. de Souza Doering, Chains of prime ideals in Noetherian rings, J. Pure Appl. Algebra, 18 (1980) 97–109.

7. A. M. de Souza Doering and Ives Lequain, The gluing of maximal ideals—applications to a problem on the spectrum of a Noetherian ring and to problems on the going up and going down in polynomial rings, Trans. Amer. Math. Soc., 260 (1980) 583–593.

8. D. Ferrand and M. Raynaud, Fibres formelles d'un anneau local Noethérien, Ann. Sci. Ecole Norm. Sup., 3 (1970) 295–311.

9. K. Fujita, Some counterexamples related to prime chains of integral domains, Hiroshima Math. J., 5 (1975) 473–485.

10. R. Gilmer, Multiplicative Ideal Theory, Marcel Dekker, New York, 1972.

11. R. Heitmann, Examples of noncatenary rings, Trans. Amer. Math. Soc., 247 (1979) 125–136.

12. M. Hochster, Prime ideal structure in commutative rings, Trans. Amer. Math. Soc., 142 (1969) 43–60.

13. E. G. Houston and S. McAdam, Rank in Noetherian rings, J. Algebra, 37 (1975) 64–73.

14. I. Kaplansky, Adjacent prime ideals, J. Algebra, 20 (1972) 94–97.

15. ______, Commutative Rings, University of Chicago Press, Chicago, Ill., 1974.

16. W. Krull, Primidealketten in allgemeinen Ringbereichen, S. B. Heidelberger Akad. Wiss. Math.-Natur. Kl., 7 (1928) 1–14.

17. ______, Idealtheorie, Chelsea, New York, 1948.

18. ______, Zum Dimensionsbegriff der Idealtheorie (Beiträge zur Arithmetik kommutativer Integritätsbereiche, III), Math. Z., 42 (1937) 745–766.

19. W. J. Lewis, The spectrum of a ring as a partially ordered set, J. Algebra, 25 (1973) 419–434.
20. H. Matsumura, Commutative Algebra, W. A. Benjamin, New York, 1970.
21. S. McAdam, Saturated chains in Noetherian rings, Indiana Univ. Math. J., 23 (1974) 719–728.
22. S. McAdam and L. J. Ratliff, Jr., Semi-local taut rings, Indiana Univ. Math. J., 26 (1977) 73–79.
23. ______, Asymptotic prime divisors and going down, 35-page preprint.
24. M. Nagata, On the chain problem of prime ideals, Nagoya Math. J., 10 (1956) 51–64.
25. ______, Note on a chain condition for prime ideals, Mem. Coll. Sci., Univ. Kyoto, 32 (1959–60) 85–90.
26. ______, Local Rings, Interscience Tracts 13, Interscience, New York, 1962.
27. E. Noether, Idealtheorie in Ringbereichen, Math. Ann., 83 (1921) 24–66.
28. ______, Eliminationstheorie und allgemeine Idealtheorie, Math. Ann., 90 (1923) 229–261.
29. D. G. Northcott, Ideal Theory, Cambridge Tracts 42, Cambridge University Press, Cambridge, 1953.
30. D. G. Northcott and D. Rees, Reductions of ideals in local rings, Proc. Cambridge Philos. Soc., 50 (1954) 145–158.
31. T. Ogoma, Non-catenary pseudo-geometric normal rings, Japan J. Math. (to appear).
32. L. J. Ratliff, Jr., On quasi-unmixed local domains, the altitude formula, and the chain condition for prime ideals (I), Amer. J. Math., 91 (1969) 508–528.
33. ______, On quasi-unmixed local domains, the altitude formula, and the chain condition for prime ideals (II), Amer. J. Math., 92 (1970) 99–144.
34. ______, Characterizations of catenary rings, Amer. J. Math., 93 (1971) 1070–1108.
35. ______, Catenary rings and the altitude formula, Amer. J. Math., 94 (1972) 458–466.
36. ______, Locally quasi-unmixed Noetherian rings and ideals of the principal class, Pacific J. Math., 52 (1974) 185–205.
37. ______, Going-between rings and contractions of saturated chains of prime ideals, Rocky Mountain J. Math., 7 (1977) 777–787.
38. ______, Chain Conjectures in Ring Theory, Lecture Notes in Math. No. 647, Springer-Verlag, New York, 1978.
39. ______, $A(X)$ and GB-Noetherian rings, Rocky Mountain J. Math., 9 (1979) 337–353.
40. ______, Notes on essentially powers filterations, Michigan Math. J., 26 (1979) 313–324.
41. ______, Two theorems on the prime divisors of zero in completions of local domains, Pacific J. Math., 81 (1979) 537–545.
42. ______, Note on asymptotic prime divisors, analytic spreads and the altitude formula, Proc. Amer. Math. Soc. (to appear).
43. ______, Integrally closed ideals and asymptotic prime divisors, Pacific J. Math. (to appear).
44. ______, Notes on Macaulay rings and the altitude formula, in preparation.
45. ______, GB-rings and polynomial rings, in preparation.
46. ______, W. Rückert, Zum Eliminationsproblem der Potenzreihenideals, Math. Ann., 107 (1933) 259–281.
47. H. Seydi, Anneaux Henséliens et conditions de chaines, Bull Soc. Math. France, 98 (1970) 9–31.
48. B. L. van der Waerden, Zur algebraischen Geometrie I, Math. Ann., 108 (1933) 113–125.
49. M. Wichman, The width of a module, Canad. J. Math., 22 (1970) 102–115.
50. M. Yoshida, A theorem on Zariski rings, Canad. J. Math., 8 (1956) 3–4.
51. O. Zariski and P. Samuel, Commutative Algebra, vol. 1, Van Nostrand, New York, 1958.
52. ______, Commutative Algebra, vol. 2, Van Nostrand, New York, 1960.

MISCELLANEA

52. So, What's New? A practice has sprung up of late (encouraged by demands for premature knowledge in certain examinations) of hurrying young students into the manipulation of the machinery of the Differential and Integral Calculus before they have grasped the preliminary notions of a *Limit* and of an *Infinite Series*.... Besides being to a large extent an educational sham, this course is a sin against the spirit of mathematical progress.

—From the Preface of *Algebra: An Elementary Text-Book for the Higher Classes of Secondary Schools and for Colleges*, by G. Chrystal, Edinburgh, 1889. (Suggested by Bertram Ross.)

PÓLYA'S COUNTING THEOREM VIA TENSORS

RUSSELL MERRIS
Department of Mathematics, California State University, Hayward, CA 94542

1. Introduction. Some of the most difficult problems in mathematics involve counting. There are several reasons for this difficulty, some of which are technical and others more conceptual. A frequently encountered technical difficulty is that the objects to be counted may not be sequentially arranged (e.g., the school bus packed with jelly beans). A common conceptual difficulty occurs when different objects are identified (considered to be equal) for enumeration purposes (e.g., it may only be of interest to know how many colors of jelly beans are in the bus). In modern language, there may be an equivalence relation imposed on the objects. The problem then is to enumerate equivalence classes.

Consider the following situation. Suppose D is a set of m objects which, for simplification, we take to be $\{1,2,\dots,m\}$. Suppose further that we wish to color the objects in D and that we have at our disposal a set C of k colors, $C=\{c_1,c_2,\dots,c_k\}$. We may think of a coloring of D as a function $f\colon D\to\{c_1,c_2,\dots,c_k\}$. Let C^m denote the set of all such colorings.

So far, so good. But now we introduce the confusion of equivalent colorings. Suppose G is a group of permutations on D. We say that two colorings f_1 and f_2 of D are *equivalent* $(\operatorname{mod} G)$ if there is a permutation $\sigma\in G$ such that $f_1\sigma=f_2$. This equivalence relation imposes a partition on the set of colorings of D. The equivalence classes so obtained are called *color patterns*. The question then becomes: How many color patterns are there? A very general and elegant theorem due to G. Pólya[1] supplies the answer.

In order to state Pólya's result, we must first define the *cycle index* of a permutation group. Recall that a permutation can be (uniquely) factored as a product of disjoint cycles. For $\sigma\in G$, let $n_1(\sigma)$ be the number (possibly zero) of cycles of length one in the disjoint cycle factorization of σ. Let $n_2(\sigma)$ be the number of cycles of length two, and so on. Next, suppose $x_1,x_2,\dots,x_m$ are m independent indeterminates. The cycle index[2] of the permutation group G is the polynomial

$$Z_G(x_1,x_2,\dots,x_m)=\frac{1}{o(G)}\sum_{\sigma\in G}\prod_{i=1}^{m}x_i^{n_i(\sigma)},$$

where $o(G)$ is the cardinality of G.

Example 1. Suppose $m=4$ and $G=\{(1)(2)(3)(4),(12)(3)(4),(1)(2)(34),(12)(34)\}$. Then $Z_G(x_1,x_2,x_3,x_4)=(x_1^4+2x_1^2x_2+x_2^2)/4$.

Example 2. Suppose $m=4$ and $G=\{(1)(2)(3)(4),(12)(34),(13)(24),(14)(23)\}$. Then $Z_G(x_1,x_2,x_3,x_4)=(x_1^4+3x_2^2)/4$.

The groups in Examples 1 and 2 are isomorphic, but they do not have the same cycle index.[3] It follows that the combinatorial structure of the permutation representation of a group is vital to the computation of the cycle index.

We can now state the special case of Pólya's theorem which answers the question posed above.

THEOREM 1. *Let D be a set of m objects, C a set of k colors, and G a group of permutations*

Russell Merris is Professor of Mathematics at California State University, Hayward. He obtained his Ph.D. in Mathematics from the University of California at Santa Barbara in 1969. He was a Fulbright Lecturer at the University of Islamabad in Pakistan in 1973, a Visiting Professor at the Instituto de Fisica e Matemática in Lisbon during 1973–74, a Professeur Invité at the Université Laval in Quebec during the summer of 1977, and a National Academy of Sciences Exchange Scientist in Prague in 1979. His research interests include multilinear algebra, combinatorics, matrix theory, and group representation theory.—*Editors*

acting on D. The number of color patterns associated with this situation is

$$Z_G(k,k,\ldots,k)=\frac{1}{o(G)}\sum_{\sigma\in G}k^{n(\sigma)},$$

where $n(\sigma)=n_1(\sigma)+n_2(\sigma)+\cdots+n_m(\sigma)$, the total number of cycles (including cycles of length one) in the disjoint cycle factorization of σ.

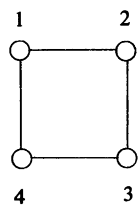


FIG. 1

Example 3. Let D be the vertices of the square in Fig. 1. Suppose we wish to color each of the vertices either red, white, or blue. Two colorings are equivalent if one can be obtained from the other by a simple rotation of the square. In the context of Theorem 1, $m=4$, $k=3$, $C=\{r,w,b\}$ and $G=\{(1)(2)(3)(4),(1234),(13)(24),(1432)\}$. For example, the colorings represented in Fig. 2 form one color pattern. With respect to the Fig. 1 numbering of vertices, the corresponding functions are: Fig. 2.1, $f(1)=r, f(2)=r, f(3)=b, f(4)=w$; Figure 2.2, $g(1)=w, g(2)=r, g(3)=r, g(4)=b$; etc. Notice that $f\sigma=g$, where $\sigma=(1432)$. Finally, observe that $h(1)=r$, $h(2)=r$, $h(3)=w$, $h(4)=b$ is in a different pattern.

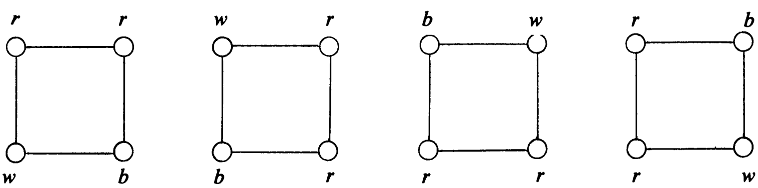


FIG. 2.1 FIG. 2.2 FIG. 2.3 FIG. 2.4

FIG. 2

How many color patterns are there? Using Theorem 1 we compute $n(1)(2)(3)(4)=4$, $n(1234)=1$, $n(12)(34)=2$ and $n(1432)=1$. Therefore, the desired number is

$$\begin{aligned}Z_G(3,3,3,3)&=\frac{1}{4}\sum_{\sigma\in G}3^{n(\sigma)}\\&=\frac{1}{4}(3^4+3+3^2+3)\\&=96/4=24.\end{aligned}$$

Indeed, this is the correct number. A list of distinct representatives for the color patterns is given in Fig. 3. (For ease of presentation, the square has been omitted.)

Example 4. Suppose we again color the vertices of the square with $C=\{r,w,b\}$, but this time let G be the group of all symmetries of the square. Referring again to Fig. 1, we find that G is now $\{(1)(2)(3)(4),(1234),(13)(24),(1432),(13)(2)(4),(1)(24)(3),(12)(34),(14)(23)\}$. The number of color patterns is

$$
\begin{aligned}
Z_G(3,3,3,3) &= \frac{1}{8}\sum_{\sigma\in G} 3^{n(\sigma)}\\
&= \frac{1}{8}(3^4+3+3^2+3+3^3+3^3+3^2+3^2)\\
&= 168/8 = 21.
\end{aligned}
$$

The braces in Fig. 3 show which Example 3 color patterns identify in the formation of Example 4 color patterns.

$$
\begin{array}{ccc}
\begin{array}{cc} r & r\\ r & r\end{array} & \begin{array}{cc} w & w\\ w & w\end{array} & \begin{array}{cc} b & b\\ b & b\end{array}\\[1em]
\begin{array}{cc} r & w\\ w & w\end{array} & \begin{array}{cc} w & r\\ r & r\end{array} & \begin{array}{cc} b & w\\ w & w\end{array}\\[1em]
\begin{array}{cc} r & b\\ b & b\end{array} & \begin{array}{cc} w & b\\ b & b\end{array} & \begin{array}{cc} b & r\\ r & r\end{array}\\[1em]
\begin{array}{cc} r & r\\ w & w\end{array} & \begin{array}{cc} w & w\\ b & b\end{array} & \begin{array}{cc} b & b\\ r & r\end{array}\\[1em]
\left\{\begin{array}{cc} r & r\\ w & b\\ r & r\\ b & w\end{array}\right. & \left\{\begin{array}{cc} w & w\\ r & b\\ w & w\\ b & r\end{array}\right. & \left\{\begin{array}{cc} b & b\\ r & w\\ b & b\\ w & r\end{array}\right.\\[1em]
\begin{array}{cc} r & w\\ w & r\end{array} & \begin{array}{cc} w & b\\ b & w\end{array} & \begin{array}{cc} b & r\\ r & b\end{array}\\[1em]
\begin{array}{cc} r & w\\ b & r\end{array} & \begin{array}{cc} w & b\\ r & w\end{array} & \begin{array}{cc} b & r\\ w & b\end{array}
\end{array}
$$

FIG. 3.

2. Proof. As suggested above, Theorem 1 is a special case of Pólya's Counting Theorem. In turn, Pólya's Theorem can be made to depend on a very beautiful result which has come to be known as Burnside's Lemma.[4] In contrast, what follows is a somewhat long and contrived proof. Why? Because the author perceives it as a device for painlessly introducing both the notion of tensor space and the connection between enumeration theory and symmetry classes of tensors.

To simplify the discussion, we consider only the case $m=4$, $k=3$, $c_1=r$, $c_2=w$, and $c_3=b$. Consider the set B of all four letter (nonsense) words which can be made up using only the letters r,w,b with repetitions allowed, i.e.,

$$B=\{a_1a_2a_3a_4\colon a_i\in\{r,w,b\},\ 1\leqslant i\leqslant 4\}.$$

Then B contains $3^4=81$ elements, some of which follow: *rrrr*, *rrrw*, *rrrb*, *rrww*, *rrwb*, *rrbw*, *rrbb*,... (The set B can be identified with the Cartesian product $C\times C\times C\times C$. For the moment, at least, we wish to play down this connection.)

The next step is to create a vector space with the set B as a basis. This can be done simply by taking all formal linear combinations of the elements of B over some field F. We will use the following somewhat cryptic notation for this vector space: $\overset{4}{\otimes} V$. The "4" comes from the fact that the words in our basis have four letters. The "$\otimes$" may be taken to suggest multiplication.

The "V" must remain momentarily enigmatic. Of course, the dimension of $\overset{4}{\otimes} V$ is 81.

It turns out that the group G acts on $\overset{4}{\otimes} V$ in a natural way. For each $\sigma \in G$, define a linear operator $P(\sigma^{-1})$ on $\overset{4}{\otimes} V$ as follows: For each word $a_1a_2a_3a_4$ in the basis, define

$$P(\sigma^{-1})a_1a_2a_3a_4 = a_{\sigma(1)}a_{\sigma(2)}a_{\sigma(3)}a_{\sigma(4)}.$$

For example, if $\sigma = (123)(4)$, then $P(\sigma^{-1})(rwbw) = wbrw$. Thus, the effect of $P(\sigma^{-1})$ is to permute the order in which the letters appear in each word. The overall effect is to permute the basis. We now extend $P(\sigma^{-1})$ linearly so that it becomes a linear operator on $\overset{4}{\otimes} V$. Finally, if $\sigma \in G$, then $\sigma^{-1} \in G$ and $(\sigma^{-1})^{-1} = \sigma$ so that $P(\sigma)$ is also defined. (One exercise which the reader may wish to undertake in order to become familiar with $\{P(\sigma) : \sigma \in G\}$ is to prove that $P(\sigma\pi) = P(\sigma)P(\pi)$ for all $\sigma, \pi \in G$.)

We require one more definition. Consider the set

$$V^4(G) = \left\{ t \in \overset{4}{\otimes} V : P(\sigma)t = t \text{ for all } \sigma \in G \right\}.$$

Observe that the "t" used here may not be an element of B, the original basis of $\overset{4}{\otimes} V$. The remainder of the proof consists of becoming familiar with $V^4(G)$. We achieve this through a sequence of lemmas.

LEMMA 1. *The set $V^4(G)$ is a subspace of $\overset{4}{\otimes} V$.*

Proof. Let $t_1, t_2 \in V^4(G)$; $d_1, d_2 \in F$; and $\sigma \in G$. Then

$$\begin{aligned} P(\sigma)(d_1t_1 + d_2t_2) &= d_1P(\sigma)t_1 + d_2P(\sigma)t_2 \\ &= d_1t_1 + d_2t_2. \end{aligned}$$

LEMMA 2. *An element $t \in \overset{4}{\otimes} V$ is in $V^4(G)$ if and only if*

$$\frac{1}{o(G)} \sum_{\sigma \in G} P(\sigma)t = t. \tag{1}$$

(One might paraphrase this by saying $P(\sigma)t = t$ for all $\sigma \in G$ if and only if $P(\sigma)t = t$ on the average.)

Proof. If $t \in V^4(G)$, then $P(\sigma)t = t$ for all $\sigma \in G$. Thus

$$\begin{aligned} \frac{1}{o(G)} \sum_{\sigma \in G} P(\sigma)t &= \frac{1}{o(G)} \sum_{\sigma \in G} t \\ &= t. \end{aligned}$$

Conversely, assume (1) holds for some $t \in \overset{4}{\otimes} V$. Let $\pi \in G$ be fixed but arbitrary. Observe

$$\begin{aligned} P(\pi)t &= P(\pi)\left(\frac{1}{o(G)} \sum_{\sigma \in G} P(\sigma)t \right) \\ &= \frac{1}{o(G)} \sum_{\sigma \in G} P(\pi)P(\sigma)t \\ &= \frac{1}{o(G)} \sum_{\sigma \in G} P(\pi\sigma)t \end{aligned}$$

$$= \frac{1}{o(G)} \sum_{\mu \in G} P(\mu)t$$

$$= t.$$

We now write

$$T = \frac{1}{o(G)} \sum_{\sigma \in G} P(\sigma).$$

As a linear combination of linear operators, T is a linear operator.

LEMMA 3. *The operator T is idempotent, i.e., $T^2 = T$.*

Proof.

$$T^2 = \frac{1}{o(G)^2} \sum_{\pi, \sigma \in G} P(\pi\sigma)$$

$$= \frac{1}{o(G)^2} \sum_{\pi \in G} \sum_{\mu \in G} P(\mu)$$

$$= \frac{1}{o(G)} \sum_{\mu \in G} P(\mu)$$

$$= T.$$

It follows from Lemmas 2 and 3 that T is the projection onto $V^4(G)$. In particular, we may conclude immediately that the dimension of $V^4(G)$ is the trace of T.

LEMMA 4. *We have*

$$\begin{aligned} \dim V^4(G) &= \operatorname{trace}(T) \\ &= Z_G(3,3,3,3) \\ &= \frac{1}{o(G)} \sum_{\sigma \in G} 3^{n(\sigma)}. \end{aligned}$$

Proof. Only the equation

$$\operatorname{trace}(T) = \frac{1}{o(G)} \sum_{\sigma \in G} 3^{n(\sigma)}$$

remains to be proved. Observe that

$$\operatorname{trace}(T) = \frac{1}{o(G)} \sum_{\sigma \in G} \operatorname{trace}(P(\sigma)).$$

Think of a matrix representation of $P(\sigma)$ with respect to the basis B of $\overset{4}{\otimes} V$. Call it $R(\sigma)$. Then $R(\sigma)$ is an 81 by 81 matrix consisting of zeros and ones. The trace of $R(\sigma)$ is the number of ones which lie on the main diagonal. A one will lie on the main diagonal in the row (and column) corresponding to the basis element $a_1a_2a_3a_4$ if and only if $a_{\sigma(1)}a_{\sigma(2)}a_{\sigma(3)}a_{\sigma(4)} = a_1a_2a_3a_4$ where $a_i \in \{r, w, b\}$, $1 \leqslant i \leqslant 4$. This happens if and only if for each cycle of σ the positions (subscripts) corresponding to that cycle all hold the same symbol. Since there are $n(\sigma)$ cycles and we can fill the positions corresponding to each cycle in 3 ways (r, w, or b), there are $3^{n(\sigma)}$ ways to do it, i.e., there are exactly $3^{n(\sigma)}$ basis elements which contribute ones to the main diagonal. Therefore, trace $(P(\sigma)) = 3^{n(\sigma)}$ and the proof is complete.

Many results in enumerative combinatorial analysis involve looking at the same set in two different ways. Such is the case here too. It is not difficult, for example, to recognize $\overset{4}{\otimes} V$ as the

vector space generated by all colorings of D. After all, what is the set B but the colorings of D? The word "$rrwb$" is a code for the function $f(1)=r$, $f(2)=r$, $f(3)=w$, and $f(4)=b$. Indeed,

$$B=\{f(1)f(2)f(3)f(4): f\in C^4\},$$

where $C=\{r,w,b\}$ and C^4 is the set of all colorings of $\{1,2,3,4\}$.

Now that $\overset{4}{\otimes} V$ has been exposed, what about $V^4(G)$? It turns out to be the vector space of color patterns in the following sense.

LEMMA 5. *Let $P_1, P_2, \ldots, P_N$ denote the color patterns into which C^4 is partitioned by the action of G. Then*

$$\left\{\sum_{f\in P_i} f(1)f(2)f(3)f(4): 1\leqslant i\leqslant N\right\} \tag{2}$$

is a basis of $V^4(G)$.

It is an immediate consequence of this lemma that N=the number of color patterns in the dimension of $V^4(G)$, and Theorem 1 is proved.

Proof. Since $\{f(1)f(2)f(3)f(4): f\in C^4\}$ is a basis of $\overset{4}{\otimes} V$ and T is the projection onto $V^4(G)$, it follows that $\{T(h(1)h(2)h(3)h(4)): h\in C^4\}$ spans $V^4(G)$. Now, let $h\in C^4$, say $h\in P_j$. Let $G_h=\{\sigma\in G: h\sigma=h\}$. It is easy to verify that G_h is a subgroup of G. Let $\sigma_1, \sigma_2, \ldots, \sigma_K$, $K=[G: G_h]$ be the distinct right coset representatives of G_h in G. Then

$$\begin{aligned}
T(h(1)h(2)h(3)h(4)) &= \frac{1}{o(G)}\sum_{\sigma\in G} h\sigma(1)h\sigma(2)h\sigma(3)h\sigma(4)\\
&= \frac{1}{o(G)}\sum_{i=1}^{K}\sum_{\sigma\in G_h} h\sigma\sigma_i(1)h\sigma\sigma_i(2)h\sigma\sigma_i(3)h\sigma\sigma_i(4)\\
&= \frac{1}{o(G)}\sum_{i=1}^{K}\sum_{\sigma\in G_h} h\sigma_i(1)h\sigma_i(2)h\sigma_i(3)h\sigma_i(4)\\
&= \frac{1}{K}\sum_{i=1}^{K} h\sigma_i(1)h\sigma_i(2)h\sigma_i(3)h\sigma_i(4)\\
&= \frac{1}{K}\sum_{f\in P_j} f(1)f(2)f(3)f(4).
\end{aligned}$$

Therefore, the set (2) spans $V^4(G)$. To prove that (2) is linearly independent, suppose

$$\begin{aligned}
0 &= \sum_{i=1}^{N} a_i \sum_{f\in P_i} f(1)f(2)f(3)f(4)\\
&= \sum_{i=1}^{N}\sum_{f\in P_i} a_i f(1)f(2)f(3)f(4).
\end{aligned}$$

But, this is precisely a linear combination of the basis of $\overset{4}{\otimes} V$. It follows that each coefficient is zero, i.e., that $a_i=0$, $1\leqslant i\leqslant N$.

3. Concluding Remarks. Earlier, we recognized $\overset{4}{\otimes} V$ in terms of colorings. In fact, it has still another name. It is (at least embryonically) the fourth tensor power of a three-dimensional

vector space V. The subspace $V^4(G)$ is the symmetry class of tensors corresponding to the group G and the principal (trivial) irreducible character. The function

$$S_G(y_1,y_2,\ldots,y_k) = Z_G\left(\sum_{i=1}^{k} y_i, \sum_{i=1}^{k} y_i^2, \ldots, \sum_{i=1}^{k} y_i^m\right)$$

is the Schur function associated with G and the principal character. Indeed, S_G is itself the character of a certain representation of the full linear group of nonsingular k by k matrices. These concepts have interesting and important generalizations involving irreducible characters of G other than the principal character.

The preparation of this article was supported in part by National Science Foundation Grant no. MCS 77-28437.

Notes

1. Pólya's theory of counting appeared in *Acta Math.*, 68 (1937) 145–254. An excellent exposition of it can be found in N. G. de Bruijn's article for *Applied Combinatorial Mathematics* (E. Beckenbach, ed.), Wiley, 1964. It turns out that much of Pólya's theory was anticipated a decade earlier by J. H. Redfield (*Amer. J. Math.*, 49 (1927) 433–455).
2. No doubt in order to honor Pólya, de Bruijn used the symbol "P" for the cycle index polynomial. Pólya himself used the symbol "Z". The idea for the cycle index polynomial predates both Pólya and Redfield. It was employed by Frobenius in his determination of the irreducible characters of the symmetric group.
3. In his 1937 paper (p. 176), Pólya gives an example of two nonisomorphic permutation groups with identical cycle indexes.
4. Indeed, Pólya's Theorem is a special case of Burnside's Lemma.

WHAT IS A COMPUTER PROGRAM?

MICHAEL GEMIGNANI

Department of Mathematical Sciences, Indiana University-Purdue University, Indianapolis, IN 46205

No doubt virtually all the people who read this sentence believe that they already know what a computer program is. The purpose of this paper is to convince you that the question is not so easily answered as you may imagine. Nor is the question without practical value. For, at least in the area of law, how courts, legislative bodies, and administrative agencies view computer programs can have broad ramifications for computer specialists, attorneys, and the general public.

The dispute concerning what legal protection may be afforded software has raged since the mid-1960's with no resolution yet in sight. Can a program be copyrighted? If so, what aspects of the program will the copyright protect? Can a program be patented? How can the Patent Office determine if a given program is sufficiently novel to warrant a patent, given the fact that some million new programs are written each year?

Another area of the law in which the notion of "program" is important is "computer crime." If you deliberately transfer a program belonging to someone else to your own files by means of a "trapdoor" or "Trojan horse" technique, have you thereby committed theft? Can one be said to

---

Professor Gemignani obtained a Ph.D. from Notre Dame under the direction of Robert Clay in point-set topology and foundations of geometry. He has held positions at St. Mary's College, Notre Dame, SUNY at Buffalo, and Smith College and is currently Chairman of Mathematical Sciences at Indiana University–Purdue University at Indianapolis. He also has a J.D. from Indiana University–Indianapolis Law School. He is the author of seven textbooks, including *Elementary Topology and Axiomatic Geometry*. He also has been Chairman of the Northeast and the Indiana Sections of the MAA. He is interested in the law relating to both higher education and computing.—*Editors*

steal electronic impulses or magnetic patterns found inside a computer, particularly if one does not disturb the original patterns but merely "copies" them? This would not be a crime according to nineteenth-century common law principles, even if the program thus appropriated had a commercial value of millions of dollars. A question even more fundamental than "What is a computer program?" is "What is a computer?" Senator Ribicoff in his Federal Computer Systems Protection Act of 1979 defines a computer as "an electronic device which performs logical, arithmetic, and memory functions by the manipulations of electronic or magnetic impulses." This definition is both too broad, because it encompasses such things as digital watches, and too narrow, because it excludes fluidic computers—for example, because they are not electronic. The message here is that members of the computer community and those involved with making laws must come together to help one another deal with the legal problems created by today's rapidly developing technology.

What then is a computer program? For simplicity we restrict the question to an applications program, a program written to solve some particular problem. Once the problem to be solved is clearly formulated, a means of solution must be outlined; thus, an algorithm is constructed. The algorithm is then set forth in that stylized diagram known as a flow chart. The flow chart is then used as a guide for encoding the algorithm, usually in a "high level" computer language such as Basic or Fortran; the algorithm expressed in such a language is known as a source program. The source program is fed into the computer by means of an input device, possibly after being transferred onto a medium such as cards or magnetic tape. The source program is "translated" by a compiler into machine language, and thus becomes an object program. The object program actuates the setting of switches inside the computer, which enables the computer to perform the underlying algorithm and solve the original problem. The program has undergone a striking metamorphosis from the initial search for a solution to a problem to a pattern of switch settings inside a machine. At what stage do we have *the* program? Is there some reality underlying all of the steps in the chain? This, however, is not the question this paper addresses; rather, "program" will be taken to mean "source program" unless some specific indication to the contrary is given.

A survey of articles which touch upon the subject yield a variety of definitions for "program." Many of these definitions involve the term "set" or "series of instructions." Usually, those who use this kind of definition seem to be using "instruction" in its ordinary sense rather than its more technical meaning of one line of source code. Obviously, defining a program to be a set of lines of source code tells us virtually nothing about what a program really is; hence, instruction must be interpreted more broadly.

Those who understand the source language in which a program is written can, at least theoretically, manually carry out the procedure the program embodies. A program may thus be viewed as an instruction manual which tells those who know how to read it how to carry out a particular task. This view, however, has several serious flaws. Programs can be hideously long and intricate. American Airlines' Sabre program for booking airplane reservations, for example, contains more than a million instructions and, if written in the form of a standard size book, would be more than 14,000 pages long. It is hard to conceive of such a program as a training manual for airline reservations clerks. Generally, too, programs are written because the problems to be solved require techniques too time-consuming and tedious to be done manually. Why then would one have an instruction book to tell a human being how to do a task that cannot practically be carried out by a human being in the first place? Furthermore, many programs require "documentation," which is itself a form of instruction manual that explains the program and how to use it properly. If the program were an instruction manual in the usual sense, we would then have an instruction manual that explained how to use another instruction manual—a possibility, but not a very elegant one.

But if a program is a set of instructions, and these instructions are not really intended for human beings, whom or what does the program instruct? One view is that the program instructs the computer. The ability to learn, to be instructed, is an attribute normally reserved to human

beings or other high forms of life. Is a computer really learning while in the process of being programmed? The problem is complicated here because the instructions of a program often look much the same as the sort of instructions human beings are accustomed to give and receive. But the computer does not, indeed cannot, see what the human being sees. The word ADD appearing in a program seems to the human reader to be a clear instruction to compute a sum, but relative to the computer it is "perceived" only as a series of switch settings. In other words, the program is merely a means to manipulate various components of the machine. The program may thus be viewed as a process or means of bringing about a desired result inside a computer.

The notion of "control" is sometimes used instead of "instruct" in defining the function of a program. If a program controls a computer in the same way that a distributor controls the sequence of firing the sparkplugs in an internal combustion engine, then the program can reasonably be viewed as an integral part of the computer itself. This view is strengthened by the fact that a general purpose computer will not, indeed cannot, carry out its appointed task until it has been properly programmed. The programming sets the switches, in effect redesigns the internal structure of the machine, becomes an inseparable part of the machine, if the machine is to perform as the program was written to make it perform. The program may thus be viewed as a machine part or as the completion of a previously incomplete machine.

The view of a program as a machine part has been promoted in the past by the Court of Customs and Patent Appeals, which has been fighting an uphill battle to have at least certain types of programs declared patentable by the United States Supreme Court. This view has been criticized, however, by Otho Ross [**8**] on the grounds that the ephemeral electrical charges that represent the program inside the computer do not really modify the hardware; the program can be modified by the user or the computer itself, even while the program is running; and in time-sharing computers, programs of many different users are constantly being "swapped" in and out of the central processor. Ross asks, "Does the computer become a 'new machine' every time the operation shifts to a new program?"

But because the program considered here is the source program, which remains outside the machine and separate from it at all times, one may object that even though the program may result in the creation of a machine part, it itself is not a machine part. Just as a person who understands the language can use a program to reconstruct its algorithm, so too such a person who also has an adequate background in electrical engineering could use the program to build an electronic circuit which could carry out the program in conjunction with input and output devices. That is, the program is, to a suitably trained engineer, a circuit diagram or the blueprint for building special-purpose hardware. Even if it is not part of a machine, the program can be used to construct a machine of a specific design, that design being contained implicitly in the program itself. The program thus stands midway between the abstract solution to a problem and a machine that actually carries that solution out.

There are still other ways to view a program. The Association for Computing Machinery defines a program to be "an ordered set of data that are coded instructions or statements that when executed by the computer cause the computer to process data." This takes the program to be a special kind of data compilation. This view becomes more comprehensible when one remembers that a program is stored in a computer as data and that the object program, when written out, appears as just a string of 0's and 1's. If "cipher" is taken in its cryptological sense, the ACM definition might inspire a view of a program as an enciphered or encrypted message.

Are there still other ways to view a program? Indeed there are. A program may be thought of as bearing the same relationship to a computer as a phonograph record does to a record player. Nevertheless, there are highly significant differences between a record and record player, on the one hand, and a program committed to a medium such as tape and a computer, on the other. First, a record is specifically intended to communicate with human beings; the program is intended to communicate with a machine. Second, the record player has as its sole purpose to reproduce the "information" contained on the record. The computer, on the other hand, uses the

program to process data. It is the data processed by the computer acting in accordance with instructions contained in the program which is more analogous to the phonograph record. But the purpose of the computer is generally not to play back the data, but to transform it. The better a record player reproduces the sound of the recording, the better player it is considered to be. A computer which merely plays back the data fed into it is of little use to anyone. Nor does the record determine what the phonograph can do; the phonograph had the ability to play the record even before it actually does so. But a computer has the ability to do very little until the program infuses it with a purpose and the means to carry that purpose out.

Finally, a program may be viewed as simply a pattern of symbols, a nonsense writing, so to speak. This view is an easy way out, avoiding any disturbing technical and philosophical questions which force human beings to look inside themselves at the very nature of their minds, or to meditate on the uneasy relationship which they have established with machines, machines which seem to mirror ever more closely, or even surpass, the workings of their brains. Perhaps herein lies one of the fundamental psychological and philosophical problems encountered in any attempt at analyzing the nature of computers or computer programs: The more deeply we analyze computers, the more deeply we must analyze ourselves, our nature as human beings. As John Hersey phrased it in his eloquent dissent contained in the Final Report of the National Commission on New Technological Uses of Copyrighted Works (CONTU):

> As one step in its education, this Commission has had the benefit of a book written by ... Professor Joseph Weizenbaum of M.I.T., entitled *Computer Power and Human Reason*—a work which is both intricately technical and profoundly humanistic. Something that Professor Weizenbaum keeps emphasizing over and over again is the extent to which computer scientists, especially those who have worked with so-called artificial intelligence—"and large segments of the general public as well"—have come to accept the propositions "that men and computers are merely two different species of a more abstract genus called 'information processing systems,'" that reason is nothing more than logic, and "that life is what is computable and only that."

Whether, or to what extent, we agree with this statement may determine our individual responses to the question asked in the title of this paper.

To add to the complexity of the question, computer programs come in a multiplicity of languages and are embodied in a variety of media. Two programs which ostensibly do the same thing can be totally different in appearance. Even two such programs written in the same source language and employing the same underlying algorithm can be strikingly different. What then is a computer program? A particular form of expression of a flowchart or algorithm? An instruction manual for human beings? A process for controlling or bringing about a desired result inside a computer? A machine part or completion of an incomplete machine? A circuit diagram or blueprint for a circuit board? A compilation of data? A coded writing? A "phonograph record," or perhaps "sheet music?" A mere pattern of symbols or a nonsense writing? All of these things, some of them, or something altogether different? When you read the first sentence of this article, you were probably sure you knew what a computer program is. Perhaps now the answer is not so obvious.

References

1. Banzhaff, Copyright protection for computer programs, Columbia Law Review, 64 (1964) 1274.
2. Bender, Computer programs: Should they be patentable?, Columbia Law Review, 68 (1968) 241.
3. ______, Trade secret protection of software, Geo. Washington Law Review, 38 (1970) 909.
4. Breyer, The uneasy case for copyright: A study of copyright in books, photocopies, and computer programs, Harvard Law Review, 84 (1970) 281.
5. Davis, Computer programs and subject matter patentability, Rutgers Journal of Computers & the Law, 6 (1977) 1.
6. Gemignani, Legal protection for computer software: The view from '79, Rutgers Journal of Computers & the Law (to appear).
7. National Commission on New Technological Uses of Copyrighted Works, Final Report, 1978.
8. Ross, The patentability of software and firmware, Computer Law Service, Section 4-2, Art. 5, 1978.

THE NINTH U.S.A. MATHEMATICAL OLYMPIAD

SAMUEL L. GREITZER
Department of Mathematics, Rutgers University, New Brunswick, NJ 08903

The Ninth U.S.A. Mathematical Olympiad was held on May 6, 1980. There were sufficient differences this time to warrant a more detailed report than usual. First, the number of contestants was increased. By setting the "cut-off" score on the Annual High School Mathematics Examination to 107, we were able to invite 121 students instead of the usual 100. In addition, five more students were invited to participate—three on the basis of their previous record involving the Olympiad and the Training Session, and two because of strong recommendations by the regional contest chairmen for the High School Math. Examination. This would have given us 126 papers to grade; but five students did not respond to the invitations and one did not return his answers, so we finally sat down to grade 120 sets of papers on Friday, May 16, at Rutgers University. Once again we were fortunate in having the help of a dedicated and interested group of faculty members. They included Professors Richard Cohn, Michael Aissen, John Bender, Richard Bumby, Solomon Leader, Ben Muckenhoupt, Hy Zimmerberg, all of Rutgers, and Professors Murray Klamkin, of the University of Alberta, and Harry Ruderman, of Hunter College High School.

Because the students are talented, the problems are usually solved in many different ways, and grading is not easy. However, by the middle of the day, all papers had been graded. No paper has the name of the contestant on it—only a number—so there is no way that any paper can receive special attention.

The same afternoon and evening, the papers were all collated, still by number, and Murray Klamkin reread the top thirty papers or so. Finally, that evening all grading and recording was complete, and the eight students who had scored highest were called and notified of their success. This is necessary because there is very little time between the grading and the ceremony to which all the finalists are invited. The finalists were:

Michael Larsen	Lexington High School	Lexington, Mass.
Eric Carlson	Munster High School	Munster, Ind.
Michael Finn	Lake Braddock Secondary	Burke, Va.
Bruce Smith	Terra Linda High School	San Rafael, Cal.
Jeremy Primer	Columbia High School	Maplewood, N.J.
Paul Feldman	Stuyvesant High School	New York, N.Y.
Daniel Scales	Westwood High School	Westwood, Mass.
David Ash	Fort William Collegiate	Thunder Bay, Ontario

It should be noted that, for the first time, a student from Canada placed among the top eight finalists. Of the five students invited to participate on recommendation, three did quite well. They placed third, fourth, and eleventh. Those recommended by regional chairmen did not do as well.

The eight finalists were invited to a ceremony in Washington, D.C., which was sponsored by International Business Machines. At a gathering in the National Academy of Sciences, the students received prizes and gifts from the National Council of Teachers of Mathematics, Mu Alpha Theta, the M.A.A., and I.B.M., as well as a sophisticated calculator from Hewlett-Packard and books donated by Academic Press; Freeman & Company; McGraw-Hill; Prindle, Weber & Schmidt; Springer Verlag; and John Wiley & Sons. There was also an excellent lecture delivered by Professor R. Creighton Buck.

After this, the contestants, parents, and guests repaired to the State Department where the banquet honoring the finalists took place amid beautiful surroundings. I would say that everybody enjoyed the entire affair. The third unusual occurrence was the announcement that

there was no International Mathematical Olympiad scheduled for 1980, which was, I am sure, a great disappointment to the five finalists who were due to be graduated in June.

At this point, for the entertainment of the many who always ask for them, we present the five problems, prepared, as usual, by the committee headed by Murray Klamkin:

NINTH U.S.A. MATHEMATICAL OLYMPIAD—MAY 6, 1980

1. A two-pan balance is inaccurate since its balance arms are of different lengths and its pans are of different weights. Three objects of different weights A, B and C are each weighed separately. When placed on the left-hand pan, they are balanced by weights A_1, B_1, C_1, respectively. When A and B are placed on the right-hand pan, they are balanced by A_2 and B_2, respectively. Determine the true weight of C in terms of A_1, B_1, C_1, A_2 and B_2.

2. Determine the maximum number of different three-term arithmetic progressions which can be chosen from a sequence of n real numbers $a_1 < a_2 < \cdots < a_n$.

3. Let

$$F_r = x^r \sin(rA) + y^r \sin(rB) + z^r \sin(rC),$$

where x, y, z, A, B, C are real and $A+B+C$ is an integral multiple of π. Prove that if $F_1 = F_2 = 0$, then $F_r = 0$ for all positive integral r.

4. The inscribed sphere of a given tetrahedron touches each of the four faces of the tetrahedron at their respective centroids. Prove that the tetrahedron is regular.

5. If $1 > a, b, c > 0$, prove that

$$\frac{a}{b+c+1} + \frac{b}{c+a+1} + \frac{c}{a+b+1} + (1-a)(1-b)(1-c) < 1.$$

Table 1 shows how the contestants scored on each of the problems.

TABLE 1

	Problem				
Score	1	2	3	4	5
21 +		1	2	2	2
16–20	28	26	8	19	4
11–15	13	11	1	6	2
6–10	27	29	2	5	0
1–5	24	11	13	19	11
0	28	42	94	69	101

To attempt to draw any conclusions from this table is to become involved in lots of controversy. However, I would guess that the major difficulty in Problem 1 was that students could not perform the algebraic manipulations needed to derive the final answer. In fact, it appeared that 28 students could not set up the initial equations. Could this be a sign of Physics Anxiety?

Problem 3 can be solved in several ways. That one-third of those who attempted it could not get anywhere is an indication that our talented students should get lots more trigonometry. In Problem 4, we were quite surprised to find that students did not know that tangents to a sphere from an external point are equal. As for Problem 5, let us be charitable and say that students did not have enough time.

Table 2 shows the total scores attained by the contestants.

TABLE 2

Score						
0	1–20	21–40	41–60	61–80	81–100	101 +
10	46	40	16	5	2	1

Solutions to these problems as well as to problems in previous Olympiads may be obtained from Professor Walter E. Mientka, 917 Oldfather Hall, University of Nebraska, Lincoln, NE 68588—at 50 cents a pamphlet.

Finally, my thanks to all who helped make this a successful Olympiad—the Regional Chairmen for getting their results to me, Murray Klamkin for preparing the Olympiad and regrading, the Rutgers faculty members and Harry Ruderman for hours of grading papers, and Professors Mientka and Shell for their help.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

A PROOF OF LUSIN'S THEOREM

MARCUS B. FELDMAN
Department of Mathematics, Washington University, St. Louis, MO 63130

The purpose of this note is to give a simple, direct proof of Lusin's theorem:

THEOREM. *Let f be a real-valued (Lebesgue) measurable function defined on a measurable set $E\subset\Re$. Given $\varepsilon>0$, there exists a set $F_\varepsilon\subset\Re$, of Lebesgue measure $|F_\varepsilon|<\varepsilon$, such that $g=f|_{E-F_\varepsilon}$ is continuous.*

The proof we give seems to be simpler and more direct than the ones we have seen in the literature. Its principal tool is the property that if A is measurable and if $\delta>0$ there exists an open set $U\supset A$ such that $|U-A|<\delta$.

Let us choose a countable basis $\{U_j\}$ of open sets for $\Re$. For each $j=1,2,3,\ldots$, let U^j be an open set including $f^{-1}(U_j)$ such that $|U^j-f^{-1}(U_j)|<\varepsilon/2^j$. If

$$F_\varepsilon=\bigcup_{j=1}^{\infty}\{U^j-f^{-1}(U_j)\}$$

then, clearly, $|F_\varepsilon|<\varepsilon$. We claim that $g=f|_{E-F_\varepsilon}$ is continuous. To see this, let us assume, for the moment, the equality

$$g^{-1}(U_j)=U^j\cap[E-F_\varepsilon]. \tag{1}$$

If U is open in $\Re$ then there exists a set M of positive integers such that $U=\cup_{j\in M}U_j$. Thus, from (1),

$$g^{-1}(U)=\left(\bigcup_{j\in M}U^j\right)\cap[E-F_\varepsilon] \text{ is open in } E-F_\varepsilon.$$

Equality (1) is easily established. Obviously,

$$g^{-1}(U_j)\subset U^j\cap(E-F_\varepsilon).$$

On the other hand,

$$U^j\cap[E-F_\varepsilon]\subset U^j\cap\left[E-\left\{U^j-f^{-1}(U_j)\right\}\right]=U^j\cap E\cap\left\{U^j-f^{-1}(U_j)\right\}^c$$

$$= U^j \cap E \cap \{U^{j^c} \cup f^{-1}(U_j)\}$$
$$= E \cap U^j \cap f^{-1}(U_j)$$
$$= f^{-1}(U_j).$$

Intersecting the first and last sets with $E-F_\varepsilon$ we obtain

$$U^j \cap (E-F_\varepsilon) \subset g^{-1}(U_j).$$

This establishes (1).

REMARK. The proof is valid if the range of f is a second countable topological space and the domain is a topological space endowed with a measure μ such that, given a measurable set A and $\delta>0$, there exists an open set U covering A such that

$$\mu(U-A)<\delta.$$

If, in addition, the domain of f has the property that any measurable set contains an approximating closed set, then f can be further restricted to such an approximating closed set on which it is continuous. This corresponds to the statement (often seen in the literature) of Lusin's theorem where $E-F_\varepsilon$ is required to be closed.

Acknowledgment. I am grateful to Professor Guido Weiss, with whom I discussed this presentation and who made many valuable suggestions.

A PROOF OF THE ARITHMETIC MEAN-GEOMETRIC MEAN INEQUALITY

ADOLF CUSMARIU
School of Mathematics, University of Minnesota, Minneapolis, MN 55455

The classical theorem of the means states that for every sequence of m positive numbers $\{x_i\}$

$$A_m = \frac{1}{m}\sum_{i=1}^{m} x_i \geqslant \prod_{i=1}^{m} x_i^{1/m} = G_m \tag{AG}$$

with equality just in case the sequence is constant.

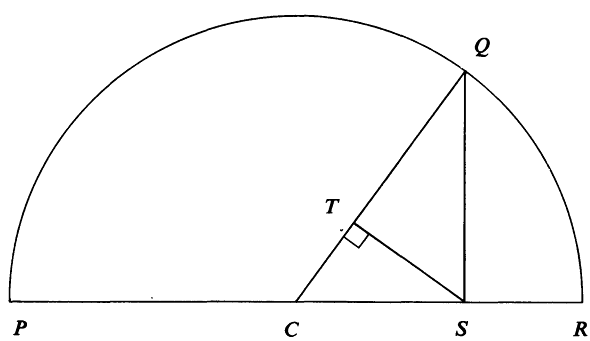


FIG. 1

Many proofs of this fundamental inequality have been discovered in modern times ([**1**], [**2**], [**3**]), but in its simplest setting—for only two numbers—the ancients possessed a singularly neat picture (see Pappus [**4**, Book 3, p. 51]), which makes matters readily apparent (see Fig. 1). Here, $\overline{CQ}$ is the arithmetic mean of the two numbers (segments) $\overline{SP}$ and $\overline{SR}$, $\overline{SQ}$ their geometric mean, and, in addition, $\overline{QT}$ is their harmonic mean.

Is there an analogous picture for three or more numbers?

During an unsuccessful attempt to obtain one, the following iterative procedure suggested itself:

$$x_1+x_2+x_3=\tfrac{1}{2}(x_1+x_2)+\tfrac{1}{2}(x_2+x_3)+\tfrac{1}{2}(x_1+x_3)\geqslant(x_1x_2)^{1/2}$$
$$+(x_2x_3)^{1/2}+(x_1x_3)^{1/2}\geqslant(x_1x_2^2x_3)^{1/4}+(x_1x_2x_3^2)^{1/4}+(x_1^2x_2x_3)^{1/4}\geqslant\cdots.$$

That is, write the first sum cyclically as a sum of averages, apply (AG) for two numbers, and then repeat. After several iterations (and some inspired guessing) the pattern of the resulting monotone sequence became clear: the exponents in the kth term are the entries of the kth power of a certain matrix, namely,

$$V=\tfrac{1}{2}(I+S)=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}&0\\0&\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&0&\frac{1}{2}\end{bmatrix}.$$

Here I is the 3×3 identity matrix and S the (shift) permutation matrix $S=(\delta_{i+1,j})$. Since V is doubly stochastic, the fundamental theorem for transition matrices (see, for example, [5, Theorem 4.1.4]) implies that $\lim_{n\to\infty}V^n=\frac{1}{3}J$, where J is the matrix of all 1's, so that if we let

$$f_n=f(X,V^n)=\tfrac{1}{3}\sum_{i=1}^{3}\prod_{j=1}^{3}x_j^{(V^n)_{ij}}$$

then $f_0=A_3$ (as usual $V^0=I$), $\lim_{n\to\infty}f_n=G_3$, and (AG) for three numbers obtains as soon as we can show formally that f_n is nonincreasing. Now, the action of the (left) shift S on any matrix is just a cyclic permutation of the rows; so $f(X,V^n)=f(X,SV^n)$. Averaging, then, amounts to nothing more than writing f_n as

$$f_n=\tfrac{1}{2}(f(X,V^n)+f(X,SV^n));$$

that is,

$$f_n=\tfrac{1}{3}\sum_{i=1}^{3}\tfrac{1}{2}\left(\prod_{j=1}^{3}x_j^{(V^n)_{ij}}+\prod_{j=1}^{3}x_j^{(SV^n)_{ij}}\right).$$

Consequently,

$$f_n\geqslant\tfrac{1}{3}\sum_{i=1}^{3}\prod_{j=1}^{3}x_j^{(\frac{1}{2}(I+S)V^n)_{ij}}=f(X,V^{n+1})=f_{n+1}$$

by an application of (AG) for two numbers and the recurrence

$$\tfrac{1}{2}(I+S)V^n=VV^n=V^{n+1}.$$

As for the matter of equality, if $A_3=f_0=\lim_{n\to\infty}f_n=G_3$, then $f_0=f_1$ because f_n is monotone; so

$$\tfrac{1}{2}(x_1+x_2)+\tfrac{1}{2}(x_2+x_3)+\tfrac{1}{2}(x_1+x_3)=(x_1x_2)^{1/2}+(x_2x_3)^{1/2}+(x_1x_3)^{1/2},$$

i.e., $x_1=x_2=x_3$.

To obtain a proof of the (AG) inequality for m numbers, all that need be done is to take the matrix V to be of size m, thereby replacing 3 by m everywhere, and the above argument goes through entirely unchanged.

It is interesting to note that the convergence (V is now $m\times m$) $\lim_{n\to\infty}V^n=(1/m)J$ may also be derived directly from the binomial theorem. Indeed,

$$V^n=2^{-n}(I+S)^n=2^{-n}\sum_{q=0}^{n}\binom{n}{q}S^q;$$

but, since $S^m = I$, the change of variables $q = tm + p$, where $0 \leqslant p \leqslant m-1$ and $0 \leqslant t \leqslant [(n-p)/m]$, shows that $S^q = S^p$. Therefore,

$$V^n = \sum_{p=0}^{m-1} \left[2^{-n} \sum_{t=0}^{\left[\frac{n-p}{m}\right]} \binom{n}{tm+p} \right] S^p;$$

so we will have the desired limit once it is clear that

$$\lim_{n\to\infty} 2^{-n} \sum_{t=0}^{\left[\frac{n-p}{m}\right]} \binom{n}{tm+p} = \frac{1}{m}.$$

For this we again borrow from the past—a combinatorial identity dating back to 1834 ([**6**]; see also [**7**, pp. 19–20] and [**8**, pp. 40–41]):

$$2^{-n} \sum_{t=0}^{\left[\frac{n-p}{m}\right]} \binom{n}{tm+p} = \frac{1}{m}\left(1 + \sum_{k=1}^{m-1} \left(\cos\frac{k}{m}\pi\right)^n \cos\frac{k}{m}\pi(n-2p)\right)$$

in which we have only to pass to the limit.

Finally, there is no reason our iterative procedure ought not work in reverse, and it does. We have

$$x_1x_2x_3 = (x_1x_2)^{1/2}(x_2x_3)^{1/2}(x_1x_3)^{1/2} \leqslant \tfrac{1}{2}(x_1+x_2)\tfrac{1}{2}(x_2+x_3)\tfrac{1}{2}(x_1+x_3)$$
$$\leqslant \tfrac{1}{4}(x_1+2x_2+x_3)\tfrac{1}{4}(x_1+x_2+2x_3)\tfrac{1}{4}(2x_1+x_2+x_3) \leqslant \cdots \uparrow \left(\tfrac{1}{3}(x_1+x_2+x_3)\right)^3.$$

More generally, if we let

$$g_n = g(X, V^n) = \left(\prod_{i=1}^{m} \sum_{j=1}^{m} (V^n)_{ij} x_j \right)^{1/m} = \prod_{i=1}^{m} (V^n X)_i{}^{1/m},$$

then, this time, $g_0 = G_m$, $\lim_{n\to\infty} g_n = A_m$, and, for the proper monotonicity,

$$g_n = (g(X, V^n) g(X, SV^n))^{1/2} \leqslant g\left(X, \tfrac{1}{2}(I+S)V^n\right) = g_{n+1}.$$

Furthermore, running the iterative procedure from either mean, with the appropriate averaging, yields induction schemes as well, both forward and backward, and, with a bit of care, proofs of the inequality with arbitrary weights may be gotten along similar lines.

Acknowledgment. I would like to thank Professor Max Jodeit for his interest and for some useful suggestions.

References

1. G. H. Hardy, J. E. Littlewood and G. Pólya, Inequalities, 2nd ed., Cambridge University Press, London, 1964.
2. E. F. Beckenbach and R. Bellman, Inequalities, 3rd ed., Ergebnisse der Mathematik, Band 30, Springer-Verlag, Berlin and New York, 1971.
3. D. S. Mitrinović, Analytic Inequalities, Die Grundlehren der Math. Wissenschaften, Band 165, Springer-Verlag, Berlin and New York, 1970.
4. Pappus of Alexandria, La Collection Mathématique, French translation by Paul Ver Ecke, vol. 1; Desclée de Brouwer et Cie, Paris, 1933.
5. J. G. Kemeny and J. L. Snell, Finite Markov Chains, Van Nostrand Reinhold, New York, 1960.
6. C. Ramus, Solution générale d'un problème d'analyse combinatoire, J. Reine Angew. Math., 11 (1834) 353–355.
7. E. Netto, Lehrbuch der Combinatorik, 2nd ed., 1927; reprinted by Chelsea, New York.
8. J. Riordan, An Introduction to Combinatorial Analysis, Wiley, New York, 1958.

ON A VANISHING PRODUCT IN THE INTEGRAL GROUP RING

S. M. GAGOLA, JR., S. C. GARRISON, AND M. R. PETTET
Department of Mathematics, Texas A & M University, College Station, TX 77843

In the Research Problems section of this MONTHLY, R. M. Robinson [2] referred briefly to some conjectures about the integral group ring $\mathbb{Z}[G]$ of a finite abelian group G that arise from certain problems in geometry (see [3]). The specific question that he posed is answered by the corollary below.

If G is written multiplicatively, $\mathbb{Z}[G]$ is the ring of $\mathbb{Z}$-linear combinations of elements of G with componentwise addition and with multiplication defined using the group multiplication and the distributive law.

THEOREM. *Suppose $g_1, g_2, \ldots, g_n$ are elements of a finite multiplicative abelian group G such that $\prod_{i=1}^{n}(1-g_i)=0$ in $\mathbb{Z}[G]$. If $|g_i|$ denotes the order of g_i and if $n>1$, then*

$$\sum_{i=1}^{n} 1/|g_i|>1.$$

Proof. Let $\hat{G}$ be the dual group of G, that is, the group of homomorphisms of G into the multiplicative group of complex numbers (under pointwise multiplication). The following facts (and their proofs) are analogous to the familiar relations between vector spaces and their duals. $\hat{G}$ is isomorphic to G since Theorems 4.6 and 10.7 of [4] reduce the question to the case in which G is cyclic, where it is easily checked. For any subgroup A of G, let A° be the subgroup of $\hat{G}$ consisting of those homomorphisms whose kernels contain A. The restriction map $\lambda \mapsto \lambda|_A$ from $\hat{G}$ to $\hat{A}$ is an epimorphism, by Theorem 10.6 of [4], and the kernel of this map is, by definition, A°. Hence, the First Isomorphism Theorem (2.12 of [4]) implies that $\hat{G}/A^\circ \simeq \hat{A}$. In particular, $|G|/|A|=|\hat{G}|/|\hat{A}|=|A^\circ|$.

Under the hypothesis of the theorem, we first prove that $\hat{G}=\cup_{i=1}^{n}\langle g_i\rangle^\circ$, where $\langle a, b,\ldots, c\rangle$ denotes the smallest subgroup of G containing $a, b,\ldots, c$. Let $\lambda \in \hat{G}$. Since λ extends by linearity to a ring homomorphism λ^* from $\mathbb{Z}[G]$ to the complex numbers, the hypothesis implies that

$$\prod_{i=1}^{n}(1-\lambda(g_i))=\lambda^*\left(\prod_{i=1}^{n}(1-g_i)\right)=0.$$

Hence, for some i, $1-\lambda(g_i)=0$, so that $\lambda \in \langle g_i\rangle^\circ$. The fact that $\hat{G}=\cup_{i=1}^{n}\langle g_i\rangle^\circ$ follows.

Properties of the dual group proved above now yield that

$$|G|=|\hat{G}|=\left|\bigcup_{i=1}^{n}\langle g_i\rangle^\circ\right|<\sum_{i=1}^{n}|\langle g_i\rangle^\circ|=\sum_{i=1}^{n}|G|/|\langle g_i\rangle|.$$

Here the strict inequality holds, as the identity element of $\hat{G}$ is counted n times in the right-hand sum. Since $|g_i|=|\langle g_i\rangle|$ by definition, the proof is completed by dividing both sides of the inequality by $|G|$.

COROLLARY. *Under the hypotheses of the theorem, $|g_i|<n$ for some i.*

In fact, it is easily proved that $\prod_{i=1}^{n}(1-g_i)=0$ if and only if $\hat{G}=\cup_{i=1}^{n}\langle g_i\rangle^\circ$, which is equivalent to the combinatorial statement

$$1-\sum_{i} 1/|g_i|+\sum_{i<j} 1/|\langle g_i, g_j\rangle|-\cdots=0.$$

Note. Professor Robinson has informed the authors that he has received similar solutions of this problem from A. W. Hales, from K. H. Kim and F. W. Roush, and also from M. Kiyota and K. Nomura. Kim and Roush have proved a refinement of the corollary, showing that $|g_i| < n-1$ for some i if $n-1$ is composite [1]. In addition, I. M. Isaacs has communicated to us a proof of the fact that, if $\prod_{i=1}^{n}(1-g_i)=0$ but no proper subproduct vanishes, then, after suitable renumbering,

$$|\langle g_1, g_2, \dots, g_t \rangle| < n(n-1)\cdots(n-t+1) \quad \text{for all } t,\ 1 \leqslant t \leqslant n.$$

References

1. K. H. Kim and F. W. Roush, Robinson's conjecture on abelian groups, preprint.
2. Raphael M. Robinson, Solutions of an equation in abelian groups, this MONTHLY, 86 (1979) 690.
3. ______, Multiple tilings of n-dimensional space by unit cubes, Math. Z., 166 (1979) 225–264.
4. Joseph J. Rotman, The Theory of Groups, An Introduction, Allyn and Bacon, Boston, 1965.

AN AFFINE GENERALIZATION OF THE EULER LINE

ERNST SNAPPER
Department of Mathematics, Dartmouth College, Hanover, NH 03755

Let ABC be a triangle in the Euclidean plane. We denote the centroid of ABC by G, its circumcenter (center of the circumscribed circle) by O, and its orthocenter (intersection of the three altitudes) by H. It was shown by Euler (1707–1783) that G, O, and H lie on a line and that $GH = 2(OG)$. There are many proofs of this theorem, and the line OGH is called "the Euler line"; see for instance [2, pp. 17, 18].

Euler's theorem refers to the metric of the Euclidean plane, since this metric is necessary to define the points O and H. We show here that this theorem is a special case of a much more general theorem that refers only to the *affine* structure of the plane. This affine theorem may, of course, mention G, but not O or H.

THEOREM. *Let P be any point in the plane whatsoever. Let L_A be the line through the vertex A that is parallel to the line through P and the midpoint A' of the opposite side BC (Fig. 1). Let the lines L_B and L_C be constructed similarly. Then, the three lines L_A, L_B, and L_C have a point Q in common, the points G, P, and Q lie on a line, and $GQ = 2(PG)$.*

Proof. Consider the dilatation d with center G and ratio -2, and put $d(P) = Q$. Since $d(A') = A$ (Fig. 1), Q lies on the line L_A. Similarly, Q lies on L_B and L_C and the remainder of the theorem follows equally easily from the properties of d.

Proof of Euler's theorem. We do not assume it known that the three altitudes of a triangle pass through a point and choose for P the circumcenter O of ABC. Then, the line through P and A' is perpendicular to the side BC (Fig. 1), and the same holds for the other two sides of ABC. Hence the theorem above now states that the three altitudes of ABC pass through a point H, that the points G, O, and H lie on a line, and that $GH = 2(OG)$.

We observe that the affine theorem above generalizes immediately to arbitrary dimensions. Let $A_0 \dots A_n$ be a simplex in n-space and let P be an arbitrary point in n-space. Through the vertex A_i of the simplex draw the line L_i that is parallel to the line through P and the centroid A'_i of the opposite face $A_0 \dots A_{i-1}A_{i+1} \dots A_n$. Then, the $n+1$ lines $L_0, \dots, L_n$ have a point Q in common, the line through P and Q passes through the centroid G of the simplex, and $GQ = n(PG)$. The proof of this affine theorem is entirely analogous to the proof given above for the case $n = 2$. Simply use the dilatation with center G and the ratio $-n$.

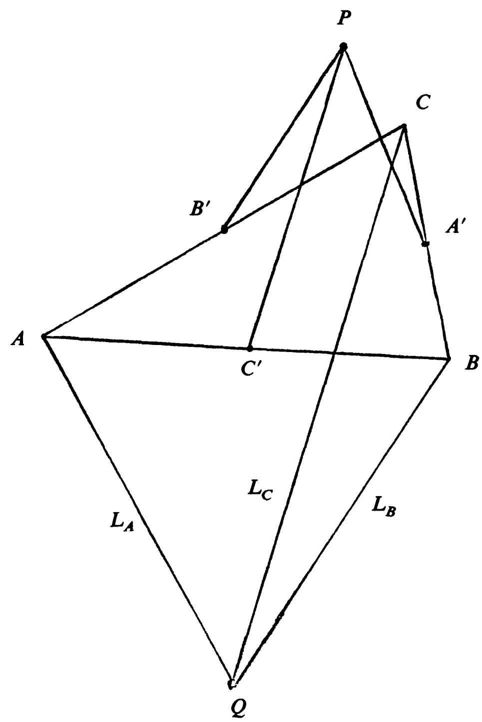


FIG. 1

However, the metric theorem, that is, Euler's theorem, does not extend to higher dimensions. When $n = 2$, A_i' is both the centroid and the circumcenter of the 1-simplex opposite to A_i and this fails in higher dimensions. There is hence the open question: What conditions should an n-simplex satisfy in order that Euler's theorem (in some reasonable form) holds for it? H. S. M. Coxeter has suggested that a clue to this riddle may be found in [1].

Comments. I. Howard Eves has observed that if one chooses for the point P the incenter I of the triangle ABC (the center of the inscribed circle of ABC), the point Q falls on the Nagel point K of the triangle. The Nagel point is the point of concurrency of the lines from the vertices of the triangle to the internal points of contact of the opposite escribed circles; see [2, p. 12], for escribed circles and [3, p. 83], for the Nagel point. Hence the points G, I, and K lie on a line and $GK = 2(IG)$.

II. Asia Weiss, a Ph.D. student of Coxeter's, has observed that the following nine points $A', B', C', A'', B'', C'', D, E, F$ lie on a "nine-point conic." (We continue the notation of the Theorem.) A', B', C' are the midpoints of the sides BC, CA, and AB, respectively; A'', B'', C'' are the midpoints of the segments AQ, BQ, and CQ, respectively; D, E, F are intersections of pairs of lines, namely, $D = L_A \cap BC$, $E = L_B \cap CA$, and $F = L_C \cap AB$. Her proof is based on the fact that the two hexagons $A'B''C'A''B'C''$ and $A'B'C'B''A''D$ are Pascal hexagons. Consequently, the six points $A', B'', C', A'', B', C''$ lie on a conic that contains the point D. Two further Pascal hexagons show that the points E and F also lie on this conic. Finally, it is not hard to see that the center N of this nine-point conic is the midpoint of the segment PQ.

The nine-point conic, observed by Asia Weiss, is clearly an affine generalization of the familiar nine-point circle of a triangle ([2, p. 18]). Precisely, if one chooses for the point P again the circumcenter of the triangle ABC, the nine-point conic becomes the nine-point circle itself. It

is indeed well known that the center of this circle is the midpoint of the segment joining the circumcenter and the orthocenter ([2, p. 71]).

If one chooses for P the incenter I of the triangle, the nine-point conic passes through the midpoints of the three sides of the triangle, the midpoints of the three segments joining the Nagel point with the vertices, and the internal points of contact of the three escribed circles. The center of this conic is the midpoint of the segment joining the incenter with the Nagel point. All this follows immediately from the comments by H. Eves and A. Weiss. It is obvious that this conic is a circle if and only if the triangle is equilateral.

Returning to the n-simplex $A_0 \dots A_n$ above, the nine points of the nine-point conic now become $3(n+1)$ points in n-space. They are the centroids of the $n+1$ faces of the simplex; the midpoints of the $n+1$ segments A_iQ; and the points of intersection of each of the $n+1$ lines L_i with its "opposite" face $A_0 \dots A_{i-1}, A_{i+1} \dots A_n$. One asks: Do these $3(n+1)$ points always lie on a quadratic hypersurface of dimension $n-1$? What can be said about this hypersurface in the case that Euler's theorem holds for the n-simplex?

An abstract of this Note has appeared in the Comptes Rendus Math. Rep. Acad. Sci. Canada, vol. 1 (1979) no. 5.

References

1. N. A. Court, Modern Pure Solid Geometry, Macmillan, New York, 1935.
2. H. S. M. Coxeter, Introduction to Geometry, Wiley, New York, 1969.
3. H. W. Eves, A Survey of Geometry, vol. 1, Allyn and Bacon, Boston, Mass. 1963.

A NOTE ON STRICT CONVEXITY

CHARLES RILEY
Mathematics Department, Keene State College, Keene, NH 03431

A norm for which $\|x+y\| = \|x\| + \|y\|$ implies 0, x, and y are collinear is said to be strictly convex (this condition implying that the midpoint of a chord joining distinct points of the unit sphere lies below the sphere). In 1936, Clarkson [1] showed that a separable Banach space may be renormed with an equivalent strictly convex norm. Using a total sequence of bounded linear functionals in $C[0, 1]$, he produces a strictly convex norm in this space and then refers to a theorem of Banach to embed B linearly and isometrically in $C[0, 1]$. Day [2] varied the proof to obtain this result in any separable normed space.

It is shown here that an elementary proof may be carried out within the given space. The argument makes no use of continuous linear functionals and is valid in any separable linear metric space.

THEOREM. *If X is a separable linear metric space, then there is an invariant linear metric d for X that yields the given topology and for which $d(x,z) = d(x,y) + d(y,z)$ only if x, y, and z are collinear.*

Proof. A paranorm p in a linear space is a function to the nonnegative reals satisfying (1) $p(0) = 0$, (2) $p(-x) = p(x)$, (3) $p(x+y) \leqslant p(x) + p(y)$, (4) if $t_n \to t$, $p(x_n - x) \to 0$, then $p(t_n x_n - tx) \to 0$. See Wilansky [3] for paranorms. Every linear metric space has its topology given by a metric d derived from a paranorm by $d(x,y) = p(x-y)$. Suppose then that p gives the topology of X and let $[x_n]$ be a countable dense set in X. Set $p_n(a) = \inf[p(a+\lambda x_n) | \lambda \in R]$ (the distance from the origin to the line through a in direction x_n). Since each p_n is a paranorm $\leqslant p$ it is easily seen that $w(a) = p(a) + \sum_{n=1}^{\infty}(1/2^n)p_n(a)$ is a paranorm in X equivalent to p. We need to show that $w(x+y) = w(x) + w(y)$ implies that 0, x, and y are collinear. This condition is equivalent to $p(x+y) = p(x) + p(y)$ and $p_n(x+y) = p_n(x) + p_n(y)$ for each n.

From $[x_n]$ select a subsequence $x_{n_i} \to x + y$. Then $p_{n_i}(x+y) \to 0$ and, since $p_{n_i}(x+y) = p_{n_i}(x) + p_{n_i}(y)$, $p_{n_i}(y) \to 0$. Thus for appropriate λ_i, $p(y + \lambda_i x_{n_i}) \to 0$. If $[\lambda_i]$ is unbounded, then $1/\lambda_{i_j} \to 0$ and $x_{n_{i_j}} = (1/\lambda_{i_j})(\lambda_{i_j} x_{n_{i_j}}) \to 0(-y) = 0$. Thus $x + y = 0$, and we have $x = y = 0$ (again using $w(x+y) = w(x) + w(y)$). Otherwise, $\lambda_i(x+y) = \lambda_i(x + y - x_{n_i}) + \lambda_i x_{n_i} \to -y$, so that if $y \neq 0$, λ_i converges, say to λ, and $\lambda_i x_{n_i} \to \lambda(x+y)$. Thus $\lambda(x+y) = -y$, and 0, x, and y are collinear.

In case p is a norm, it is easily checked that w is also.

References

1. J. A. Clarkson, Uniformly Convex Spaces, Trans. Amer. Math. Soc., 40 (1936) 396–414.
2. M. M. Day, Strict Convexity and Smoothness, Trans. Amer. Math. Soc., 78 (1955) 516–528.
3. A. Wilansky, Functional Analysis, Blaisdell, 1964, 52–63.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

IS BINARY NOTATION OPTIMAL?

KLAUS WEIHRAUCH
Fachbereich Mathematik, Fernuniversität Gesamthochschule D-5800 Hagen, West Germany

Natural numbers usually are represented in binary (or equivalently decimal) notation. We know that addition is very easy in this representation. A finite automaton with two one-way input tapes and one one-way output tape (shortly, "finite machine") is sufficient. However, there is no finite machine for multiplication (easy proof). Two binary numbers can be multiplied on a Turing machine in time $O(n \cdot \log n \cdot \log\log n)$ as Schönhage and Strassen [1] have proved. There is another representation, in which numbers can be multiplied but not added on a finite automaton, namely "prime-representation." For any number take the prime decomposition and write down in reversed binary notation the exponents of the prime numbers, separated by a special symbol. Example: 252 (decimal) has the notation 01a01aa1. However, there is no finite machine for addition in this representation.

There is a third representation in which addition as well as multiplication can be performed on a finite machine, "term-representation," defined by the rules: "0" and "1" are terms; if t_1 and t_2 are terms, then "$(t_1 + t_2)$" and "$t_1 \cdot t_2$" are terms. The represented number in binary representation can be found by "evaluation." However, this last representation also has a disadvantage. While the equivalence problem for binary notation and for prime representation can be decided on a finite machine, this is not possible in the case of term representation.

Problem. Is there a representation of the natural numbers by words such that for addition, multiplication, and deciding the equivalence problem finite machines are sufficient?

The answer should be "no," because we believe that binary notation cannot be bad. The problem certainly becomes more difficult if "finite machines" is replaced by "linear time

bounded Turing machines." But in this case also the answer should be "no." Two questions for upper bounds immediately arise in our context.

Problem. Find a good upper time bound for addition in prime representation.

Problem. Find a good upper time bound for solving the equivalence problem in term representation.

Questions of the dependence of data representation and computational complexity have been studied from a more general point of view in **[2]**.

References

1. A. Schönhage and V. Strassen, Schnelle Multiplikation grosser Zahlen, Computing, 7 (1971) 281–292.
2. R. Verbeek and K. Weihrauch, Data representation and computational complexity, Theoret. Comput. Sci., 7 (1978) 99–116.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

THE LINEAR DIOPHANTINE EQUATION

STANLEY KERTZNER
Department of Mathematics, Hofstra University, Hempstead, NY 11550

Introduction. A simple algorithm is presented for finding the general solution of the diophantine equation

$$y_1c_1+y_2c_2+\cdots+y_nc_n=e \tag{1}$$

where $c_1, c_2,\ldots,c_n$, e are given integers.

Standard methods generally employ "two-pass" procedures. In **[1]** and **[5]**, an algorithm is repeatedly used to arrive at a relation, which is then fed back into the algorithmic equations, enabling a solution to be read off. Other methods (**[6]**) rely on the ability to solve the equation when $n=2$ and then proceed to an inductive solution when $n>2$. Computer programs (**[3]**) oftentimes give a particular solution only. A strictly arithmetic method for the general solution to the equation was given by Lehmer in 1919 **[4]**, in which a "two-pass" method is utilized, at the end of which the n^2 cofactors of a determinant of order n must be computed.

The algorithm presented in this paper provides a direct, noninductive, "one-pass" procedure for finding all solutions of (1). The method is strictly arithmetic in character and is readily programmed for computer processing. As part of the procedure, the greatest common divisor, d, of $c_1, c_2,\ldots, c_n$ is found, as well as integers $y_1, y_2,\ldots, y_n$ for which $y_1c_1+y_2c_2+\cdots+y_nc_n=d$.

We present the theory first and then describe the method with an example. While the theory utilizes matrix methods, it may easily be translated into traditional number-theoretic terms.

The Theory. If M is a matrix, we agree to denote the ith row of M by M_i. We also let

$$XC=D$$

where X is an $n\times n$ integer matrix such that $\det X=\pm 1$, $C=(c_1c_2\cdots c_n)^T$ is a nonzero $n\times 1$ integer matrix, and D is an $n\times 1$ matrix such that for some integer k, $1\leqslant k\leqslant n$,

ON UNIFORM CONTINUITY AND COMPACTNESS IN METRIC SPACES

HERMANN HUEBER
Fakultät für Mathematik, Universität Bielefeld, D-4800 Bielefeld 1, West Germany

Almost every textbook on analysis or topology contains a proof of the fact that on a compact metric space every continuous real-valued function is bounded and uniformly continuous.

It would seem to be natural to ask which metric spaces E have the property that all continuous real-valued functions on E are bounded and which metric spaces E have the property that all continuous real-valued functions on E are uniformly continuous.

In 1948 Hewitt answered the first question: A metric space E is compact if and only if every continuous function from E to $\mathbb{R}$ is bounded (cf. [**1**, p. 69]).

In the present paper, we give a similar answer to the second question: In Theorem 1 we give a complete description of all metric spaces that have the property that each continuous real-valued function on E is uniformly continuous. In Theorem 2 we use this result to give a new necessary and sufficient criterion for compactness.

In what follows, E will denote a metric space with metric d. For any $x\in E$ and any subset D of E we shall denote by $d(x,D)$ the distance from x to D, that is, $d(x,D)\equiv\inf\{d(x,y)|\,y\in D\}$. By $d(x)$ we shall denote the distance from x to $E\setminus\{x\}$. Remember that a point $x\in E$ is called an accumulation point of a subset S of E if $d(x,S\setminus\{x\})=0$. Hence the set of accumulation points of E, which will be denoted by A, is just the set $\{x\in E\,|\,d(x)=0\}$. The set of isolated points is just the set $\{x\in E\,|\,d(x)>0\}$.

THEOREM 1. *With the notations formulated above, the following three conditions on E are equivalent*:

(i) *Every continuous function $f\colon E\to\mathbb{R}$ is uniformly continuous.*
(ii) *Every sequence $(x_n)_{n\in\mathbb{N}}$ in E with $\lim_{n\to\infty}d(x_n)=0$ has a convergent subsequence.*
(iii) *The set A is compact, and for every $\delta_1>0$ there exists $\delta_2>0$ such that for all $x\in E$ with $d(x,A)>\delta_1$ we have $d(x)>\delta_2$.*

Proof. (i)$\Rightarrow$(ii): Assume that (ii) is false. Hence there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in E with $\lim_{n\to\infty}d(x_n)=0$, a sequence which has no convergent subsequence. Since $\lim_{n\to\infty}d(x_n)=0$, there exists a sequence $(y_n)_{n\in\mathbb{N}}$ in E such that $\lim_{n\to\infty}d(x_n,y_n)=0$ and $y_n\neq x_n$ for every $n\in\mathbb{N}$.

Let $(y_{n_k})_{k\in\mathbb{N}}$ be a convergent subsequence of $(y_n)_{n\in\mathbb{N}}$. Since $\lim_{k\to\infty}d(x_{n_k},y_{n_k})=0$ the corresponding subsequence $(x_{n_k})_{k\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ would converge to the same limit. Therefore the sequence $(y_n)_{n\in\mathbb{N}}$ has no convergent subsequence. Hence in both sequences no point occurs infinitely often.

Now it is not difficult to see that one can choose an increasing sequence $(n_k)_{k\in\mathbb{N}}$ of natural numbers such that $X\equiv\{x_{n_k}|\,k\in\mathbb{N}\}$ and $Y\equiv\{y_{n_k}|\,k\in\mathbb{N}\}$ are disjoint sets. Since both sequences $(x_{n_k})_{k\in\mathbb{N}}$ and $(y_{n_k})_{k\in\mathbb{N}}$ have no convergent subsequence the sets X and Y are closed. By Urysohn's Lemma (cf. [**2**, p. 115]) there exists a continuous function $f\colon E\to\mathbb{R}$ that has the value one on X and the value zero on Y.

As k tends to infinity $d(x_{n_k},y_{n_k})$ tends to zero and hence $(f(x_{n_k})-f(y_{n_k}))/d(x_{n_k},y_{n_k})$ tends to infinity. So we have constructed a continuous function f that is not uniformly continuous, and therefore (i) is not fulfilled.

(ii)$\Rightarrow$(iii): If (ii) holds then every sequence in A has a convergent subsequence, and it is clear that the limit again is in A. Hence A is compact. Let $\delta_1>0$ and let $\delta_2\equiv\inf\{d(x)|\,x\in E,\,d(x,A)>\delta_1\}$. Assume $\delta_2=0$. Hence there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in E with $\lim_{n\to\infty}d(x_n)=0$ and $d(x_n,A)>\delta_1$ for every $n\in\mathbb{N}$. By (ii) there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ that converges to a point y. Obviously y lies in A. Hence $\lim_{k\to\infty}d(x_{n_k},A)=\lim_{k\to\infty}d(x_{n_k},y)=0$, which contradicts $d(x_n,A)>\delta_1$ for every $n\in\mathbb{N}$. Hence $\delta_2>0$ and thus (iii) is fulfilled.

(iii)$\Rightarrow$(i): Assume that (iii) is fulfilled. Let $f\colon E\to\mathbb{R}$ be continuous and let $\varepsilon>0$.

For every $x\in A$ let $\delta_x>0$ be such that $|f(y)-f(x)|<\frac{1}{2}\varepsilon$ for every $y\in E$ with $d(y,x)<\delta_x$. Since A is compact there exist finitely many $x_1,\dots,x_n\in A$ such that $A\subset\{y\,|\,d(x_k,y)<\frac{1}{3}\delta_{x_k}$ for

some $k \in \{1, \ldots, n\}\}$. Let $\delta_1 \equiv \frac{1}{3}\inf\{\delta_{x_1}, \ldots, \delta_{x_n}\}$ and choose $\delta_2 > 0$ according to (iii). Let $\delta \equiv \inf\{\delta_1, \delta_2\}$. Now let $x, y \in E$ such that $d(x, y) < \delta$.

If $d(x, A) > \delta_1$ then we have $d(x) > \delta_2$, and therefore $d(x, y) < \delta \leqslant \delta_2$ is possible only for $x = y$. In this case $|f(x) - f(y)| < \varepsilon$ is clear. If $d(x, A) < \delta_1$ then there exists $a \in A$ with $d(x, a) < \delta_1$. By construction we have $d(a, x_k) < \frac{1}{3}\delta_{x_k}$ for at least one $k \in \{1, \ldots, n\}$. For this k we have also $d(y, x_k) \leqslant d(y, x) + d(x, a) + d(a, x_k) < \delta + \delta_1 + \frac{1}{3}\delta_{x_k} \leqslant \delta_{x_k}$. Hence we have $|f(x) - f(y)| < |f(x) - f(x_k)| + |f(x_k) - f(y)| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon$. This shows that f is uniformly continuous.

It is notable that in the proof of (i) $\Rightarrow$ (ii) Urysohn's Lemma can be avoided: The desired function may be constructed directly with the aid of the metric and countable sums of countable infima and suprema.

THEOREM 2. *E is compact if and only if both the following hold: every continuous function $f: E \to \mathbb{R}$ is uniformly continuous, and for every $\varepsilon > 0$ the set $\{x \in E \mid d(x) > \varepsilon\}$ is finite.*

Proof. It is well known that on a compact space every continuous function is uniformly continuous, and thus the first condition is necessary to provide compactness of E. If for some $\varepsilon > 0$ the set $\{x \in E \mid d(x) > \varepsilon\}$ were infinite, then the family $(\{y \in E \mid d(x,y) < \varepsilon\})_{x \in E}$ would be an open covering of E that has no finite subcovering. This shows that also the second condition is necessary for compactness of E.

Now assume that both conditions are fulfilled, and let $(x_n)_{n \in \mathbb{N}}$ be any sequence in E. We have to show that $(x_n)_{n \in \mathbb{N}}$ has a convergent subsequence. This is trivial if some point of E occurs infinitely often in the sequence. So we may assume that no point occurs infinitely often in our sequence. Since for every $\varepsilon > 0$ the set $\{x \in E \mid d(x) > \varepsilon\}$ is finite, this implies $\lim_{n \to \infty} d(x_n) = 0$. But now (i) $\Rightarrow$ (ii) of Theorem 1 shows that there exists a convergent subsequence.

References

1. E. Hewitt, Rings of real-valued continuous functions I, Trans. Amer. Math. Soc., 64 (1948) 45–99.
2. J. L. Kelley, General Topology, Van Nostrand, Princeton, Toronto, Melbourne, London, 1955.

BOUNDS FOR THE ZEROS OF POLYNOMIALS

EMERIC DEUTSCH
Department of Mathematics, Polytechnic Institute of New York, Brooklyn, NY 11201

Consider the polynomial

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0, \tag{1}$$

where $a_0, a_1, \ldots, a_{n-1}$ are complex numbers. For every zero z of $f(z)$ we have

$$|z| \leqslant \max\{|a_0|, 1 + |a_1|, \ldots, 1 + |a_{n-1}|\}, \tag{2}$$

$$|z| \leqslant \max\{1, |a_0| + |a_1| + \cdots + |a_{n-1}|\}, \tag{3}$$

$$|z| \leqslant r, \tag{4}$$

where r is the unique positive zero of

$$g(z) = z^n - |a_{n-1}|z^{n-1} - \cdots - |a_1|z - |a_0|. \tag{5}$$

These are classical results, due basically to Cauchy (see, for example, [**1**, pp. 96–97], [**2**, p. 126]). In the present note we shall derive a series of new upper bounds for the absolute values of the zeros of f, which generalize those given by (2), (3), and (4).

PROPOSITION 1. *Every zero of the complex polynomial*

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

satisfies

$$|z| \leqslant \max\{r_k, 1+|a_{k+1}|, 1+|a_{k+2}|, \ldots, 1+|a_{n-1}|\},$$

where $k \in \{0,1,\ldots,n-1\}$ *and* r_k *is the unique positive zero of*

$$g_k(z) = z^{k+1} - |a_k|z^k - |a_{k-1}|z^{k-1} - \cdots - |a_1|z - |a_0|.$$

Proof. Let $M = \max_{j=k+1,\ldots,n-1}|a_j|$. We shall show that a complex number z satisfying $|z| > 1+M$ and $|z| > r_k$ cannot be a zero of f. We have

$$\begin{aligned} |f(z)| &\geqslant |z|^n - \left|\sum_{j=0}^{n-1} a_j z^j\right| \geqslant |z|^n - \sum_{j=0}^{n-1} |a_j||z|^j \\ &= |z|^n - |z|^{k+1} + |z|^{k+1} - \sum_{j=0}^{k} |a_j||z|^j - \sum_{j=k+1}^{n-1} |a_j||z|^j \\ &\geqslant |z|^n - |z|^{k+1} + g_k(|z|) - M(|z|^n - |z|^{k+1})/(|z|-1) \\ &= (|z|^n - |z|^{k+1})(|z|-1-M)/(|z|-1) + g_k(|z|). \end{aligned}$$

Since $|z| > r_k$ and r_k is the unique positive zero of g_k, we have $g_k(|z|) > 0$ and thus $|f(z)| > 0$.

Remark 1. It is easy to see that for $k = 0$ and $k = n-1$ Proposition 1 yields the inequalities (2) and (4), respectively.

Remark 2. Denoting by M_k the upper bound given by Proposition 1, i.e.,

$$M_k = \max\{r_k, 1 + |a_{k+1}|, 1 + |a_{k+2}|, \ldots, 1 + |a_{n-1}|\},$$

we have $M_{k+1} \leqslant M_k$ $(k = 0, 1, \ldots, n-2)$.

Indeed, applying Proposition 1 to g_{k+1}, one obtains $r_{k+1} \leqslant \max\{r_k, 1 + |a_{k+1}|\}$, which implies $M_{k+1} \leqslant M_k$. Thus, the bound given by Proposition 1 improves with the increase of k. However, this occurs at the expense of the computational difficulties in the evaluation of r_k.

Corollary. *Every zero of the complex polynomial*

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

satisfies

$$|z| \leqslant \max\{1, |a_0| + |a_1| + \cdots + |a_k|, 1 + |a_{k+1}|, \ldots, 1 + |a_{n-1}|\},$$

where $k \in \{0, 1, \ldots, n-1\}$.

Proof. This is an immediate consequence of Proposition 1 and inequality (3), since the latter applied to r_k and g_k yields $r_k \leqslant \max\{1, |a_0| + |a_1| + \cdots + |a_k|\}$.

Remark 3. For $k = 0$ and $k = n-1$, the Corollary yields the inequalities (2) and (3), respectively.

Example. Consider the polynomial

$$f(z) = z^3 + 0.3z^2 + 0.7z + 0.7.$$

Inequalities (2) and (3) give the same upper bound, namely, 1.7, while Proposition 2 with $k=1$ yields

$$|z| \leqslant \max\{|a_0| + |a_1|, 1 + |a_2|\} = 1.4.$$

References

1. M. Marden, The Geometry of the Zeros of a Polynomial in a Complex Variable, American Mathematical Society, New York, 1949.

2. M. Parodi, La Localisation des Valeurs Caractéristiques des Matrices et ses Applications, Gauthier-Villars, Paris, 1959.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

Send all proposed problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Same Enumerator for Distinct Posets

S 20 [1979, 702]. *Proposed by A. P. Hillman, University of New Mexico.*

Let n be a nonnegative integer and let S consist of all ordered quintuples $Q = (x_1, x_2, x_3, x_4, x_5)$ of nonnegative integers x_i with $x_1 + x_2 + x_3 + x_4 + x_5 = n$. Prove or disprove that there are exactly the same number of Q in S with $x_2 \leqslant x_3 \leqslant x_4 \leqslant x_5$ as there are satisfying the simultaneous conditions

$$x_1 \leqslant x_2 \leqslant x_4, \quad x_1 \leqslant x_3 \leqslant x_4, \quad \text{and} \quad x_3 \leqslant x_5.$$

Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, the Netherlands. Since $x_1 + x_2 + x_3 + x_4 + x_5 = x_1 + 4x_2 + 3(x_3 - x_2) + 2(x_4 - x_3) + (x_5 - x_4)$, one sees that the number of elements in S satisfying $x_2 \leqslant x_3 \leqslant x_4 \leqslant x_5$ equals the number of solutions of

$$y_1 + 4y_2 + 3y_3 + 2y_4 + y_5 = n, \qquad y_i \geqslant 0,$$

which in turn equals the coefficient of z^n in

$$\frac{1}{1-z} \cdot \frac{1}{1-z^4} \cdot \frac{1}{1-z^3} \cdot \frac{1}{1-z^2} \cdot \frac{1}{1-z}. \tag{1}$$

Now the simultaneous conditions

$$x_1 \leqslant x_2 \leqslant x_4, \quad x_1 \leqslant x_3 \leqslant x_4, \quad \text{and} \quad x_3 \leqslant x_5$$

are equivalent to the following 5 cases:

$$\begin{aligned}
&x_1 \leqslant x_2 \leqslant x_3 \leqslant x_4 \leqslant x_5\\
&x_1 \leqslant x_2 \leqslant x_3 \leqslant x_5 < x_4\\
&x_1 \leqslant x_3 < x_2 \leqslant x_4 \leqslant x_5\\
&x_1 \leqslant x_3 < x_2 \leqslant x_5 < x_4\\
&x_1 \leqslant x_3 \leqslant x_5 < x_2 \leqslant x_4.
\end{aligned}$$

For these five cases we use an argument similar to that above. We obtain

$$\frac{1}{1-z}\cdot\frac{1}{1-z^2}\cdot\frac{1}{1-z^3}\cdot\frac{1}{1-z^4}\cdot\frac{1}{1-z^5}\{1+z+z^3+z^4+z^2\}$$

$$=\frac{1}{1-z}\cdot\frac{1}{1-z^2}\cdot\frac{1}{1-z^3}\cdot\frac{1}{1-z^4}\cdot\frac{1}{1-z}. \tag{2}$$

Equality of (1) and (2) proves the assertion.

Also solved by George E. Andrews, D. M. Bloom, Anne L. Furno, I. P. Goulden (Canada), Richard Stanley, and the proposer.

Note. Stanley states that the following related problem is still open: If P is an n-element partially ordered set and $(a_0, a_1, a_2, \ldots)$ a sequence of nonnegative integers satisfying $\Sigma a_i = n$, then define $f_P(a_0, a_1, a_2, \ldots)$ to be the number of maps $\sigma : P \to \{0, 1, 2, \ldots\}$ such that $\sigma(x) < \sigma(y)$ whenever $x < y$ and such that exactly a_i elements x of P satisfy $\sigma(x) = i$. Do there exist two different P and Q for which $f_P(a_0, a_1, a_2, \ldots) = f_Q(a_0, a_1, a_2, \ldots)$ for all $(a_0, a_1, a_2, \ldots)$? He has checked that $n \geqslant 7$ if such P and Q exist.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by July 31, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2873. *Proposed by K. Satyanarayana, Hyderabad, India.*

Let m and n be positive integers, and

$$P_{n,m} = \sum_{k=0}^{[n/2]} (-1)^k \binom{n+1}{k} (n+1-2k)^m,$$

where $[n/2]$ denotes the greatest integer less than or equal to $n/2$. Show that

(i) $P_{n,m} = 0$ for $n = m + 1 + 2p$, $p = 0, 1, \ldots$.
(ii) $P_{n,m} > 0$ for $n = m - 1 - 2p$, $p = 0, 1, \ldots, [(m-1)/2]$.
(iii)* $P_{2n+1,2m+1} = (-1)^{n-m}\binom{2n}{n} f(n)/(2n-1)(2n-3)\cdots(2n+1-2m)$, where $f(n)$ is a polynomial (of degree m) in n with positive coefficients.
(iv)* $P_{2n,2m}$ is given by the same formula.
(v)* From (iii), (iv) it would follow that $P_{n,m} > 0$ when $m < n$, and $(-1)^{(n-m)/2} P_{n,m} > 0$ when $n \geqslant m$, provided $n + m$ is even. Show this directly.

In the above, n, m are positive integers, and $f(n)$ is a polynomial (of degree m) in n with positive coefficients.

E 2874. *Proposed by Naoki Kimura and Tetsundo Sekiguchi, University of Arkansas.*

Let $n \geqslant 3$, $0 < A_i \leqslant \pi/2$, $i = 1, 2, \ldots, n$. Assume $\Sigma_{i=1}^n \cos^2 A_i = 1$. Prove

$$\Sigma \tan A_i \geqslant (n-1)\Sigma \cot A_i.$$

E 2875. *Proposed by David Singmaster, Polytechnic of South Bank, England.*

(a) Which rational numbers m/n can be expressed as $1/x + 1/y$ where x and y are positive integers?

(b)* One can define a density on the rationals m/n with m, n positive by considering m/n as

an ordered pair (m, n). For any set A of such ordered pairs,

$$\delta(A) = \lim_{x\to\infty} |A \cap \{1, 2, \ldots, x\}^2|/x^2,$$

when the limit exists. Is the set of rationals (m,n), with no common factor, of the form $1/x + 1/y$, of density 0?

E 2876. *Proposed by Doug Hensley, Texas A & M University.*

Let $f(n) = \sum_{k=1}^{n} 1/k$, $n \geqslant 1$.

(a) Is there a continuation of f to an analytic function defined on the interval $x \geqslant 0$ such that $f(x+1) - f(x) = (x+1)^{-1}$?

(b)* Is there an analytic continuation of f to the complex half-plane $\operatorname{Re}(z) > 0$ such that $f(z+1) - f(z) = (z+1)^{-1}$?

E 2877. *Proposed by J. C. Lagarias and A. M. Odlyzko, Bell Laboratories, Murray Hill, N.J.*

Let $f(k)$ be any sequence of integers with $f(1) = 1$, $f(2) = 2$, $f(3) = 3$ and which is defined recursively by using for each $k \geqslant 4$ one of the following rules, subject to the constraint that Rule 1 can never be used twice consecutively.

Rule 1. $f(k) = f(k-1)$.

Rule 2. $f(k) = f(k-2) + f(k-3)$.

Find the largest constant θ such that $\liminf_{k\to\infty} (f(k))^{1/k} \geqslant \theta$ is valid for all such sequences $f(k)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Uniform Convergence on $(0, \infty)$

E 2784 [1979, 504]. *Proposed by F. S. Cater, Portland State University.*

For each positive integer n and each positive number x, let $F_n(x) = 0$ if $x < (n+1)^{-1}$, and let

$$F_n(x) = k^{-1}\left[(n+1)^{-1} + 2(n+2)^{-1} + \cdots + k(n+k)^{-1}\right]$$

if $x \geqslant (n+1)^{-1}$, where k is the largest integer satisfying

$$(n+1)^{-1} + (n+2)^{-1} + (n+3)^{-1} + \cdots + (n+k)^{-1} \leqslant x.$$

Let $F(x) = \lim_{n\to\infty} F_n(x)$ for $x > 0$. Determine the function $F(x)$. Is the convergence of $F_n(x)$ uniform in x? Find $\sup F(x)$ and $\sup[F(x)/x]$.

Solution by Michael Dixon, California State University, Chico. For $x > 0$ write $x = \log(1+t)$. We estimate, for fixed n and x, the integer $k = k(x, n)$ and obtain $[nt] \leqslant k \leqslant [nt+t]$. First,

$$\sum_{i=1}^{[nt]} 1/(n+i) \leqslant \int_n^{[nt]+n} du/u \leqslant \log(1+t) = x,$$

so that $[nt] \leqslant k$ follows. Similarly, setting $M = [nt+t]$,

$$\sum_{i=1}^{M+1} 1/(n+i) > \int_{n+1}^{M+n+1} du/u > \int_{n+1}^{(n+1)(1+t)} du/u = \log(1+t),$$

and hence $k \leqslant M = [nt+t]$.

For fixed n the function $F_n(x) = F_n(x, k)$ increases with k so that the upper bound for k yields

$$F_n(x) = F_n(x, k) \leqslant F_n(x, M) \leqslant \frac{1}{M}\int_1^{M+2} \frac{u}{n+u}\, du$$

$$=1-\frac{n}{M}\log\frac{n+M+2}{n+1}\leqslant 1-\frac{n}{(n+1)t}\log(1+t),$$

or

$$F_n(x)-\left(1-\frac{\log(1+t)}{t}\right)\leqslant\frac{\log(1+t)}{(n+1)t}<\frac{1}{n+1}. \tag{1}$$

The lower bound $k\geqslant[nt]$ gives

$$F_n(x)\geqslant F_n(x,[nt])\geqslant\frac{1}{[nt]}\int_0^{[nt]}\frac{u}{n+u}du$$

$$=1-\frac{n}{[nt]}\log\frac{n+[nt]}{n}\geqslant 1-\frac{n}{[nt]}\log(1+t).$$

If n is chosen so large that $nt-1>0$, then $F_n(x)\geqslant 1-n\log(1+t)/(nt-1)$, or

$$F_n(x)-\left(1-\frac{\log(1+t)}{t}\right)\geqslant\frac{-\log(1+t)}{t(nt-1)}\geqslant\frac{-1}{nt-1}. \tag{2}$$

From (1) and (2) we have

$$\left|F_n(x)-\left(1-\frac{x}{e^x-1}\right)\right|\leqslant\max\left(\frac{1}{n+1},\frac{1}{nt-1}\right),\qquad n\geqslant N(x). \tag{3}$$

Thus $F(x)=1-x/(e^x-1)$, and from (3) the convergence is clearly uniform on the interval $[\delta,\infty)$ for $\delta>0$.

Uniform convergence on $[0,\infty)$ follows from the fact that $F(x)$ is increasing, $F(0)=0$, from the inequalities $0\leqslant F_n(x)$, $F_n(x)-F(x)<1/(n+1)$, and from the uniform convergence on $[\delta,\infty)$ for each $\delta>0$.

The assertion $\sup F(x)=1$ is immediate from the series representation of e^x. Elementary calculations yield $\lim F(x)/x=\frac{1}{2}$ $(x\to 0^+)$. Moreover, $(F(x)/x)'<0$ provided

$$x[(x-1)e^x+1]<(e^x-1)[(e^x-1)-x].$$

This is equivalent to $(x/2)^2<\sinh^2(x/2)$, which is true for all $x>0$, so that $\sup F(x)/x=\frac{1}{2}$ follows.

Also solved by G. Gripenberg (Finland), Thomas Hermann (Hungary), Thomas Jager, and the proposer.

Disjoint Neighborhoods and Countable Local Bases

E 2806 [1979, 864]. *Proposed by F. S. Cater, Portland State University.*

Let S denote a topological space in which every compact set is closed, and let x and y be distinct points of S.

(1) Prove that x and y have disjoint neighborhoods if each of x and y has a countable local base.

(2) Show by example that x and y need not have disjoint neighborhoods if each element of S, other than x, has a countable local base.

Solution to (1) *by A. A. Jagers, Technische Hogeschool Twente, Enschede, Netherlands.* Let $\{A_n\colon n=1,2,\dots\}$ and $\{B_n\colon n=1,2,\dots\}$ be countable local bases at x and y, respectively. Clearly, we may assume that $A_n\supset A_{n+1}$ and $B_n\supset B_{n+1}$ for all n. Since any finite set is compact and every compact subset of S is closed, S is T_1. In particular, $\{x\}$ and $\{y\}$ are closed, and so we may assume that $y\notin A_n$ and $x\notin B_n$. Now suppose that $A_n\cap B_n$ is not empty for every n. Let $x_n\in A_n\cap B_n$. Then $\{x_n\}$ converges to both x and y, and $x_n\neq x$, $x_n\neq y$ for all n. Since $\{x_n\}$

converges to x, the set $X=\{x_n\}\cup\{x\}$ is compact. However, X is not closed since y is a limit point of X which is not contained in X. This contradiction shows that A_n and B_n are disjoint neighborhoods of x and y, respectively, from some n on.

The (counter) example for (2) given by most solvers is the one-point compactification of the rationals (or of a first countable Hausdorff not locally compact space).

Also solved by B. Brockmeyer, M. Ingenoso & B. Widynski, Chico Problem Group, A. Fora (Jordan), N. E. Frangos, C. McMullen, R. Montgomery, M. Mrševič (Yugoslavia), E. Posti (Finland), S. M. Robinson, P. S. Schnare (Saudi Arabia), A. Smuckler (Israel), J. Tripp, and the proposer.

Brockmeyer, Ingenoso, and Widynski found the problem in A. Wilansky's *Topology for Analysis*, Problem 117 (p. 51), Problem 107 (p. 86), and Problem 205 (p. 144).

Iterations Converging to a Root

E 2808 [1979, 864]. *Proposed by Peter Henrici, Swiss Federal Institute of Technology.*

Let $p(z)=a_0+a_1z+\cdots+a_kz^k$, where the a_i are complex numbers and $a_0\neq 0$. Ordinary iteration applied to p in the form

$$q_{n+1}=\frac{-a_0}{a_1+q_n(a_2+q_n(a_3+\cdots+q_na_k)\cdots)} \tag{1}$$

may or may not produce a sequence $\{q_n\}$ that converges to a zero of p. Show, however, that if (1) is replaced by

$$q_{n+1}=\frac{-a_0}{a_1+q_n(a_2+q_{n-1}(a_3+\cdots+q_{n-k+2}a_k)\cdots)} \tag{2}$$

then for almost all choices of starting values $(q_1, q_2,\dots,q_{k-1})$ the sequence $\{q_n\}$ converges to the zero of smallest modulus of p, if p has a single such zero.

Solution by L. Van Hamme, Free University of Brussels, Belgium. Let $\alpha_1, \alpha_2,\dots,\alpha_t$ be the zeros of p, with $|\alpha_1|<|\alpha_i|$ for $i\neq 1$. The recurrence relations for q_n can be written in the form

$$a_0+a_1q_{n+1}+a_2q_{n+1}q_n+\cdots+a_kq_{n+1}q_n\cdots q_{n-k+2}=0.$$

Divide by $q_1q_2\cdots q_{n+1}$ and put $U_n=(q_1q_2\cdots q_n)^{-1}$, to obtain the linear recurrence

$$a_0U_{n+1}+a_1U_n+\cdots+q_kU_{n-k+1}=0,$$

with characteristic equation $p(1/z)=0$. Hence U_n is of the form

$$U_n=c\alpha_1^{-n}+\sum_{i=2}^{t}q_i(n)\alpha_i^{-n}, \tag{*}$$

where $q_i(n)$ is a polynomial in n.

For every choice of the starting values for which $c\neq 0$, we conclude from (*) that, as $n\to\infty$,

$$\lim q_n\equiv\lim U_{n-1}/U_n=\lim c\alpha_1^{-n}/c\alpha_1^{-(n-1)}=\alpha_1.$$

Also solved by Michael Golomb, P. S. Schnare (Saudi Arabia), and the proposer.

Golomb showed that the result holds even if α_1 is a multiple zero. Schnare noted that Henrici's *Elements of Numerical Analysis* (Wiley) contains a substantial part of the information given in the proof above. The iteration is a variant of "Bernoulli's method."

Primality of 2^n+n^2

E 2809 [1980, 60]. *Proposed by Leo J. Alex, SUNY, College at Oneonta.*

Let n be a positive integer greater than 1.

(a) Show that if 2^n+n^2 is a prime, then $n\equiv 3 \pmod 6$.

(b) Investigate the converse.

(a) *Solution by uncounted solvers.* Mod 6, 2^n+n^2 is 2, 5, 2, 3, 4, 3 if n is $\equiv$ 2, 3, 4, 5, 0, 1, respectively.

(b) If $n=27$, then $f(n)=2^n+n^2$ is a sum of two cubes and not a prime. Wilfred D. Costa (undergraduate, Bombay, India) chose a, b, s, r with $a\not\equiv 0 \bmod 3$, $a|2^r+1$, $3sr\equiv 1$ or $-1 \bmod a$, and noted that 2^n+n^2 is not prime if $n=3sr+6arb$. Kenneth L. Bernstein noted $f(n)$ is not prime if $n=3^{3k}$. Richard Beigel noted that, if $f(n)$ is prime and if $m>n$, then $f(m)$ is composite if $m=n+3kf(m)(2^n+n^2-1)$, $k=1, 2, \ldots$. Arnold Adelberg referred to Hardy and Wright, *Theory of Numbers*, page 18, Theorem 22. Lee Erlebach noted $f(n)$ is prime for $n=3, 9, 15, 21, 31$; he and Jeff Shallit (undergraduate) verified that $f(n)$ is otherwise composite if $n\leqslant 332$. N. Franceschine noted that $17|f(n)$ if $n=6m+3$ and $m\equiv 0, 6, 7, 10, 16, 21, 29, 43 \bmod 68$. M. Ahuja noted that $17|f(n)$ if $n=12j+3$, $j=3, 5, 9 \bmod 16$. Ken Brown noted that $521|f(n)$ if $n=6\cdot 521^j+3$. Jordi Dou (Spain), R. K. Oliver, B. C. Oltikar, Daniel Finkel, University of South Alabama Problem Group, and University of Hartford Problem Group gave similar results. F. S. Cater (among others) noted that $2, 3, 5, 7\nmid f(n)$ if $n\equiv 3 \bmod 6$. All these results depend on periodicities. A. A. Jagers (Netherlands) gave a general theorem in this direction: Let p be a prime, and let q be a primitive root of p. Let $g(n)$ be an integer sequence ($n=2, 3, \ldots$) that is periodic mod p, with period p. Let k be the number of times per period that $g(n)\equiv 0 \bmod p$. Then $f(n)=q^n+g(n)$ is periodic mod p with period $p(p-1)$, and $f(n)$ is divisible by p for exactly $p-k$ values of $n \bmod p(p-1)$. This theorem includes most of the results above. For example, $11|f(n)\Leftrightarrow n\equiv 29, 41, 45, 57, 65, 83, 91, 93$, or $97 \bmod 110$.

Sum of $(-1)^{[2^n x]}$

E 2813 [1980, 60]. *Proposed by A. D. Sands, University of Dundee, Scotland.*

For $0\leqslant x<1$, show that

$$\sum_{n=1}^{\infty}\frac{(-1)^{[2^n x]}}{2^n}=1-2x,$$

where $[u]$ is the greatest integer in u. Also, find the sum of the series for $x\geqslant 1$.

Solution by many solvers. If $0\leqslant x<1$, let $x=\Sigma a_k 2^{-k}$ be the binary representation of x. Then $(-1)^{[2^n x]}=(-1)^{a_n}$, so that the answer to the problem is $1-2x$. If $x\geqslant 1$, a similar argument gives $1-2(x-[x])$, valid also if $0\leqslant x<1$ (in fact valid for any real x).

F. S. Cater, Ronald Evans, and Mauri Orjatsalo (Finland) based their solutions on the functional equation $2S(\frac{1}{2}x)=1+S(x)$, so that $f(2x)=2f(x)$ [if $S(x)=1-2f(2x)$]. Russell Lyons proved a generalization: For any real-valued function g, and for any integer m,

$$\Sigma_{n=-m}^{\infty}\ 2^{-n}g([2^n x])=2^{m+1}g(0)+[g(1)-g(0)]x',$$

where $x'=x-2^{m+1}[2^{-m-1}x]$.

Solutions were received also from A. Adelberg, M. Ahuja, A. Bager (Denmark), D. W. Bailey, R. Beigel (student), K. L. Bernstein, A. Bešlagič (student, Yugoslavia), W. A. Beyer, T. Q. Binh (student, Hungary), D. M. Bloom, W. Boucher (student, Canada), W. G. Brady, R. Breusch, P. S. Bruckman, P. G. de Buda (Canada), R. C. Carson, J. E. Chance, J. P. Coughlin & J. F. Morrison, M. J. Dixon, M. P. Eisner, M. Feinstein, L. L. Foster, N. Franceschine III, B. Gallassio & M. Watkins (students), P. R. Gardner, C. T. Giel, N. Glick, M. Golomb, R. P. Grimaldi, V. Hernandez (Spain), J. D. Hiscocks, I. Jungreis, C. Kalicki, H. Kappus (Switzerland), S. A. Katre (India), P. G. Kirmser, M. F. Kruelle (student), L. Kuipers (Switzerland), P. Kumar (India), D. Leep, J. Levy, G. N. Lewis, O. P. Lossers (Netherlands), J. L. de Miguel (Spain), R. B. Nelsen, B. M. O'Connor, R. K. Oliver, E. Posti (Finland), D. Redmond, M. Rieck (Germany), K. Roberts (student, Canada), R. Rossa, O. G. Ruehr, T. Šalát (Czechoslovakia), P. S. Schnare (Saudi Arabia), J. Shallit (student), J. H. Silverman, D. A. Singer, A. Smuckler (Israel), St. Olaf College Problem Group, J. Suck (Germany), E. Triesch (student, Germany), J. Tripp, E. Trost (Switzerland), University of South Alabama Problem Group, E. R. Von Eschen, C. R. Wall, W. V. Webb, J. Young, J. Zaslow, P. J. Zwier, and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by July 31, 1981. The solver's full post-office address should be on each sheet.

6334. *Proposed by Helmut Prodinger, Technische Universität, Vienna, Austria.*

Let $\Sigma=\{a,b,\dots\}$ be an alphabet. Let Σ^* be the set of all (finite) strings or words $\{x,y,\dots\}$ that use symbols of Σ. Write $x<.y$ to mean that y can be obtained by inserting exactly one (extra) symbol somewhere into the string x. The generalized binomial symbol $\binom{x}{z}$ is the number of ways that z occurs as a subsequence of x. [If $x=bbabab$, $z=bb$, $\binom{x}{z}=\binom{4}{2}=6$.] Define $\binom{x}{\varepsilon}=1$, where ε is the empty word. Prove or disprove: For every $x\in\Sigma^*$, there is a sequence $\varepsilon=z_0<.z_1<.\cdots<.z_n=x$ such that the sequence of integers $\binom{x}{z_i}$ is unimodal. ($\{a_i\}$ is unimodal if, for some r, $a_0\leqslant\cdots\leqslant a_r$, and $a_r\geqslant a_{r+1}\geqslant\cdots$.)

6335. *Proposed by Harley Flanders, Florida Atlantic University.*

In the ring $R=C[X,Y,Z]$, find the primary decomposition of the ideal

$$A=(X^3-Y^2,Y^3-Z^2,Z^3-X^2).$$

6336. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let $a_1,a_2,\dots,a_k$ be real numbers, and suppose that

$$\lim_{n\to\infty}\sin(na_1)\sin(na_2)\cdots\sin(na_k)=0.$$

Prove that at least one of the a_i is an integral multiple of π.

6337. *Proposed by R. Daniel Mauldin, North Texas State University, and Jan Mycielski, University of Colorado.*

Prove that if h is a measure-preserving ergodic homeomorphism of the unit square S onto itself then there exists a set $D\subset S$ such that D is of measure 1, $S-D$ is meager and for every $p\in D$ the sequence $p,h(p),hh(p),\dots$ is dense in S.

6338. *Proposed by Robert E. Shafer, Berkeley, California.*

Show that for z,s real, $-1\leqslant s<2$, $s\neq 1$, $z\geqslant 0$,

$$\lim_{N\to\infty}\sum_{n=0}^{N}\left(z+n+\frac{1}{2}\right)^{s-2}-\frac{(z+N+1)^{s-1}}{s-1}>\frac{2}{s(1-s)}\left(z^2+\frac{1}{4}\right)^{s/2}\sin\left(s\tan^{-1}\frac{1}{2z}\right).$$

SOLUTIONS OF ADVANCED PROBLEMS

Solution to Bessel Equation

5794 [1971, 411]. *Proposed by Walter Leighton, University of Missouri.*

Consider the (modified) Bessel equation $y''+p(x)y=0$, where

$$p(x)=1+\frac{1-4n^2}{4x^2},$$

and suppose there is a nonnull solution with zeros at $x=a$ and $x=b$ (that is, b is conjugate to a), $0<a<b$. Prove that if $y(x)$ is any solution such that $y(a)\neq 0$, then the integral

$$\int_a^b [y'^2(x)-p(x)y^2(x)]\,dx$$

is positive when $n^2>\frac{1}{4}$ and negative when $n^2<\frac{1}{4}$.

Solution by James Duemmel, Western Washington University. We may assume that a and b are consecutive zeros since the general case will follow as a sum of integrals involving consecutive zeros.

Let

$$F(y)=\int_a^b [y'^2(x)-p(x)y^2(x)]\,dx.$$

An integration by parts for the first term yields, for any solution of $y''+py=0$,

$$F(y)=y(b)y'(b)-y(a)y'(a).$$

Let z be the given nonnull solution with consecutive zeros a and b. Then $F(z)=0$. We may assume $z(x)>0$ on (a,b). Let w be the solution of $y''+py=0$ such that $w(a)=1$ and $w'(a)=0$. Then z and w are linearly independent and every solution of $y''+py=0$ is of the form $y=Az+Bw$ for some constants A and B. If $y(a)\neq 0$, then $B\neq 0$, and conversely. Substituting $y=Az+Bw$ into the integral form of F yields

$$F(y)=A^2F(z)+B^2F(w)+2AB\int_a^b [z'w'-pzw]\,dx.$$

After integrating the first term in the integral by parts and using $F(z)=0$ and $w'(a)=0$ we obtain

$$F(y)=B^2F(w)=B^2w(b)w'(b).$$

Thus we must prove $w(b)w'(b)$ is positive when $n^2>\frac{1}{4}$ and negative when $n^2<\frac{1}{4}$.

Since z and w are independent solutions there must be one and only one zero c of w between a and b. Then $w(x)>0$ on $[a,c)$, $w(x)<0$ on $(c,b]$, and $w'(c)<0$. Since $w(b)$ is negative we must show $w'(b)$ is negative when $n^2>\frac{1}{4}$ and positive when $n^2<\frac{1}{4}$.

Direct calculation verifies that when y is a solution of $y''+py=0$ then $u=y'$ is a solution of

$$\left(\frac{1}{p}u'\right)'+u=0$$

on any interval in which $p(x)\neq 0$. On any interval where $p(x)>0$ the substitution $\sqrt{p}\;v=u$ changes this equation to

$$v''+qv=0$$

with

$$q=p+\frac{3(\frac{1}{4}-n^2)}{\left(x^2+(\frac{1}{4}-n^2)\right)^2}.$$

On an interval where $p(x)>0$, the v, u, and y' related as above will have the same zeros.

When $\frac{1}{4}>n^2$ it will be true that $q(x)>p(x)>0$ for all $x>0$. Using the Sturm Comparison Theorem we compare the zeros of the solution $v=(1/\sqrt{p}\,)w'$ of $v''+qv=0$ to the zeros of the solution z of $y''+py=0$; v, and therefore w', must have a zero t in the interval (a,b).

If we collect our information about w, we see that $w''=-pw$, $w>0$ on $[a,c)$, $w<0$ on $(c,b]$, and hence $w''<0$ on $[a,c)$, $w''>0$ on $(c,b]$. Thus w' decreases on $[a,c]$ and increases on $[c,b]$. With $w'(a)=0$, $w'(c)<0$; this implies the zero of w' in (a,b) must be in (c,b) and that $w'(b)>0$. This completes the case $n^2<\frac{1}{4}$.

The case $n^2>\frac{1}{4}$ would follow by a similar argument except that it is complicated by the fact

that $p(x)$ changes sign at $\alpha=\sqrt{n^2-\frac{1}{4}}$.

If $b\leqslant\alpha$, then $z''=-pz>0$ on (a,b) so that z' must increase on (a,b). Since $z'(a)>0$, this would contradict Rolle's theorem. Hence $b>\alpha$.

Suppose $a<\alpha$. Since $w(a)=1$ and $p(a)<0$, there is some $s>a$ with $w''=-pw>0$ on $[a,s)$. Since $w'(a)=0$, this implies $w'(x)>0$ on (a,s). Since $w'(c)<0$, we know $s<c$. Let s_0 be the least upper bound of such numbers s. Then $w'(x)>0$ on (a,s_0), $w'(s_0)=0$ and $s_0<c$. w increases on $[a,s_0]$ so that $w(x)\geqslant 1$ on $[a,s_0]$. If $s_0\leqslant\alpha$, then $w''=-pw>0$ on $[a,s_0)$ and hence w' increases on $[a,s_0]$. This would contradict $w'(s_0)=0$. Therefore $\alpha<s_0$.

For $x>\alpha$ we can again use the substitution $\sqrt{p}\ v=u$. When $n^2>\frac{1}{4}$ we have $q<p$. Hence between two zeros of w' on the interval $x>\alpha$ there must be a zero of z.

If $\alpha\leqslant a$, then since $w'(a)=0$ and a and b are consecutive zeros of z, $w'(x)=0$ on $(a,b]$ is not possible since this would force a zero of z in (a,b). With $w'(c)<0$ this yields $w'(b)<0$, as we wished to prove.

If $\alpha>a$, then $w'(s_0)=0$ with $\alpha<s_0<c<b$ as in the previous notation. On $(s_0,b]$ the equation $w'(x)=0$ cannot be true since this would force a zero of z in (s_0,b). With $w'(c)<0$ this yields again $w'(b)<0$.

The Differences of the Partition Function

6137* [1977, 141; 1978, 830; 1980, 495]. *Proposed by I. J. Good, Virginia Polytechnic Institute and State University.*

Let $p(n)$ denote the number of partitions of n ($n=1,2,\ldots$), and let k denote an integer greater than 3. Prove that $\Delta^k p(n)$ ($n=1,2,\ldots$) is a sequence of alternating terms.

IV. *Further comments by the proposer.* (i) In the table on page 831 of this MONTHLY, 85 (1978), $g(k)+k$ was given instead of $g(k)$, the smallest value of n_0 for which apparently $\Delta^k p(n)\geqslant 0$ whenever $n\geqslant n_0$.

(ii) Dr. R. A. Gaskins has kindly carried the calculations up to $k=30$ and I find that $f(k)$ ceases to be a good approximation to $g(k)$ for $k>14$, but $\pi k^{5/2}$ and $\pi(k-\frac{1}{24})^{5/2}$ are. For example, $g(20)=5600$, and $g(30)=15{,}386$, while $\pi(20-\frac{1}{24})^{5/2}=5591$, and $\pi(30-\frac{1}{24})^{5/2}=15{,}433$. The fraction $\frac{1}{24}$ is suggested by the Hardy-Ramanujan-Rademacher formula for $p(n)$.

Function $f(t) = t + \tan t$

6206 [1978, 282]. *Proposed by Gérard Letac, Université Paul-Sabatier, Toulouse, France.*

Prove that, if n is an integer $\neq 0$,

$$\int_{-\pi/2}^{+\pi/2}\exp[2in(x+\tan x)]\,dx=0.$$

Solution by the proposer. The map $t\mapsto t+\tan(t)$ preserves Lebesgue measure on $\mathbb{R}$ (see [1]). Introduce functions $f_n(t)$ defined on $\mathbb{R}$ by $n\pi-\pi/2<f_n(t)<n\pi+\pi/2$ and $f_n(t)+\tan(f_n(t))=t$. Since the Lebesgue measure is preserved we get $\Sigma_{-\infty}^{\infty}f_n'(t)=1$ for all t. Now $f_n'(t)=f_0'(t-n\pi)$ and $F(t)=\Sigma_{-\infty}^{\infty}f_0'(t+n\pi)=1$ for all t. Hence

$$\int_{-\pi/2}^{+\pi/2}F(t)\exp(2\,int)\,dt=0$$

if n is an integer $\neq 0$. But this implies that

$$\int_{-\pi/2}^{\pi/2}F(t)\exp(2\,int)\,dt=\int_{-\infty}^{\infty}f_0'(t)\exp(2\,int)\,dt.$$

So setting $x=f_0(t)$ in the last integral gives the result.

Reference

1. Gérard Letac, *Which functions preserve Cauchy laws?*, Proc. Amer. Math. Soc., 67 (1977) 277–286.

Also solved by Peter Addor (Switzerland), B. D. Aggarwala & C. Nasim, K. F. Anderson, S. J. Bernau, Paul R. Chernoff, L. E. Clarke (England), Columbia University Problem Group, M. L. Glasser, Gustaf Gripenberg (Finland), Dave Joyner & Fred Roush, O. P. Lossers (Netherlands, 3 solutions), John Milcetich, Nicholas Passell, Otto G. Ruehr, and Lajos Takács.

Sets of Functions of Length Less Than 2

6257 [1979, 132]. *Proposed by Jan Mycielski, University of Colorado, Boulder.*

Give a proof of the following theorem, announced without proof by S. Banach in the Annales de la Société Polonaise de Mathématique during the 1920's:

Let X be the space of continuous nondecreasing functions $f:[0,1]\to[0,1]$ having $f(0)=0$ and $f(1)=1$ and with the distance function $d(f,g)=\max|f(x)-g(x)|$ over $0\leqslant x\leqslant 1$. Let Y be the subset of all f in X such that f is strictly increasing and the length of f is 2. Prove that $X-Y$ is meager in X. (Here the length of f is the supremum of the lengths of piecewise linear functions with finitely many vertices which agree with f on all vertices.)

Solution by Charles Riley, Keene State College, Keene, New Hampshire. Clearly each f in X has length $\leqslant 2$. Let $A_n=\{f\in X\mid \text{length of } f\leqslant 2-1/n\}$, and $B(g,\varepsilon)$ be an open ball in X. Within ε of g we will find a function h close enough to a step function such that length of h is $>2-1/n$. Then for some δ each function in $B(h,\delta)$ has length $>2-1/n$, so that $B(h,\delta)\cap A_n=\varnothing$. This proves A_n is nowhere dense in X.

Now let $C=\{(r,s)|0\leqslant r<s\leqslant 1, r,s \text{ rational}\}$ and $D_{rs}=\{f\in X\,|\,f(r)=f(s)\}$. D_{rs} is nowhere dense in X, for if $g\in X$ and B is a neighborhood of g, we may find $h\in B$ such that $h\notin D_{rs}$, and with $\delta=\frac{1}{2}(h(s)-h(r))$, we see that functions within δ of h are out of D_{rs}.

Since $X-Y=\bigcup\{A_n|n\geqslant 1\}\cup\bigcup\{D_{rs}|(r,s)\in C\}$, $X-Y$ is meager in X. Since X is complete, this shows Y is nonempty.

Also solved by F. S. Cater, L. E. Clarke (England), Paul Eitner, James B. Essick, Gustaf Gripenberg (Finland), Robert B. Israel, M. D. Mavinkurve (India), Ivan Netuka (Czechoslovakia), Nicholas Passell, and Paul Perlmutter & Steve Janke.

Several solvers noted that, for f in X, f has length 2 if and only if $f'(x)=0$ almost everywhere. Netuka notes that this assertion appears in A. C. Zaanen, *Integration*, North-Holland, Amsterdam, 1967, Exercise 42.13.

The Condition $f(x,y)<g(x)g(y)$

6266 [1979, 311]. *Proposed by Leopoldo Nachbin, University of Rochester.*

It is easily shown that every countable set S has the following property:

(P) Given any function $f\colon S\times S\to R_+$, there exists a function $g\colon S\to R_+$ such that $f(x,y)\leqslant g(x)g(y)$ for all x,y in S.

It can be shown that (P) fails if the cardinal number of S is at least equal to that of the continuum. Can it be shown without the Continuum Hypothesis that (P) fails when S is uncountable?

Solution by Martin Markl, Prague. Take S uncountable. If the assertion holds for S, then the same is true for all $S'\subset S$.

We can choose an uncountable set $S'\subset S$ for which there exists a 1-1 mapping of the set S' onto an uncountable set $S''\subset(0,1)$. In the following we will suppose directly $S'\subset(0,1)$. We will construct $f\colon S'\times S'\to R_+$ for which the assertion does not hold. Let us define

$$f(x,y)=\begin{cases} 1/|x-y| & \text{for } x\neq y \\ 1 & \text{for } x=y \end{cases}$$

Suppose that there is a function $g:S'\to R_+$ such that $f(x,y)\leqslant g(x)g(y)$. Let $S_n=\{x\in S':g(x)\leqslant n\}$. It suffices now to prove that S_n is finite for all n; hence $S'=\cup S_n$ is countable, which would be a contradiction. But if S_n is not finite, we can choose an infinite sequence $\{x_k\}\in S_n, x_{k_1}\neq x_{k_2}$ for $k_1\neq k_2$. Now, there is a subsequence x_{k_p} such that $x_{k_p}\to x$, $x\in(0,1)$. But then $(x_{k_p},x_{k_{p+1}})\to(x,x)$; hence $f(x_{k_p},x_{k_{p+1}})\to\infty$. From $f(x_{k_p},x_{k_{p+1}})\leqslant g(x_{k_p})g(x_{k_{p+1}})\leqslant n^2$ we have the contradiction.

Also solved by F. S. Cater, C. S. Gardner, Gustaf Gripenberg (Finland), Mary Ellen Rudin, and the proposer.

Relative Integral Bases in Towers of Fields

6268 [1979, 398]. *Proposed by Gene Smith and Hugh M. Edgar, San Jose State University.*

Assume that the algebraic number field K possesses at least one proper intermediate field E, i.e., $Q\subset E\subset K$. Prove or disprove the following: K must have a strictly increasing chain

$$Q=K_0\subset K_1\subset K_2\subset\cdots\subset K_{n-1}\subset K_n=K,\qquad n\geqslant 2,$$

of subfields such that K_i has a relative integral basis over K_{i-1} for $1\leqslant i\leqslant n$. [Here Q is the rational number field and $S\subset T$ means that S is a proper subset of T.]

Solution by Hugh M. Edgar and Brian Peterson, San Jose State University. Theorem 2 of "Some contributions to the theory of cyclic quartic extensions of the rationals," by Hugh Edgar and Brian Peterson, *Journal of Number Theory*, 12 (1980), reads as follows: Suppose the quadratic field F is contained in some cyclic quartic extension of Q and suppose that F has even (wide) class number. Then there is a cyclic quartic extension K of Q containing F such that K has no relative integral basis over F.

Any such tower $Q\subset F\subset K$ will serve to disprove the assertion in the problem. The easiest example is probably $F=Q(\sqrt{10})$, $K=Q\left(\sqrt{10+\sqrt{10}}\right)$.

Asymptotic Behavior of Sequences Involving e^n and $n!$

6271 [1979, 509]. *Proposed by Michael Barr, McGill University.*

For positive integers n define

$$a_n=\frac{n-1}{n}+\frac{(n-1)(n-2)}{n^2}+\cdots+\frac{(n-1)!}{n^{n-1}},$$

$$b_n=\frac{n}{n+1}+\frac{n^2}{(n+1)(n+2)}+\cdots+\frac{n^{n-1}}{(n+1)\cdots(2n-1)}.$$

(A) Prove that, for all $n>1$, $0<b_n-a_n<1$.
(B*) Prove or disprove that $\lim_{n\to\infty}(b_n-a_n)=2/3$ and that $b_n-a_n-2/3=O(1/n)$.

Solution by Lajos Takács, Case Western Reserve University. We can write that

$$a_n=(p_0+p_1+\cdots+p_{n-2})/p_n$$

and

$$b_n=(p_{n+1}+p_{n+2}+\cdots+p_{2n-1})/p_n$$

for $n\geqslant 2$ where $p_j=e^{-n}n^j/j!$ for $j=0,1,2,\ldots$.

1. If $i=1,2,\ldots,n-1$, then

$$\frac{p_{n+i}}{p_{n-1-i}}=\frac{n^{2i+1}(n-1-i)!}{(n+i)!}=\prod_{r=1}^{i}\left(\frac{n^2}{n^2-r^2}\right)>1$$

or

$$p_{n-1-i}<p_{n+i}. \tag{1}$$

Summing (1) over $i=1,2,\ldots,n-1$ and dividing by p_n we get $a_n<b_n$.

2. If $i=1,2,\ldots,n$, then $p_{n+i}/p_{n-i}=d_1d_2\cdots d_i$ where $d_r=n^2/(n+r)(n-r+1)$. Since $d_1<1$ and $d_1<d_2<\cdots<d_n$, there is a $k=2,\ldots,n$ such that $p_{n+i}<p_{n-i}$ for $1\leqslant i<k$ and $p_{n+i}>p_{n-i}$ for $k<i\leqslant n$. Hence $(k-i)p_{n+i}<(k-i)p_{n-i}$ for $1\leqslant i<k<i\leqslant n$, and this implies that

$$\sum_{i=0}^{n}(k-i)p_{n+i}<\sum_{i=0}^{n}(k-i)p_{n-i}. \tag{2}$$

Obviously, we have

$$\sum_{j=0}^{\infty}p_j=1,\quad \sum_{j=0}^{\infty}jp_j=n,\quad\text{and}\quad \sum_{j=0}^{\infty}(j-n)p_j=0.$$

By the last equation $\Sigma_{j=0}^{2n}(j-n)p_j<0$ or

$$\sum_{i=0}^{n}ip_{n+i}<\sum_{i=0}^{n}ip_{n-i}. \tag{3}$$

If we add (2) and (3), we see that

$$k\sum_{i=0}^{n}p_{n+i}<k\sum_{i=0}^{n}p_{n-i}$$

and thus $b_n-a_n<1-(p_{2n}/p_n)<1$.

3. Let us define $\nu(n)$ as a discrete random variable taking on nonnegative integers only such that $P\{\nu(n)=j\}=p_j=e^{-n}n^j/j!$ for $j=0,1,2,\ldots$. Then we can write that

$$b_n-a_n=2+\frac{1-P\{\nu(n)\geqslant 2n\}-2P\{\nu(n)\leqslant n\}}{P\{\nu(n)=n\}}.$$

Here by Chebyshev's inequality

$$0\leqslant P\{\nu(n)\geqslant 2n\}=P\{2^{\nu(n)}\geqslant 2^{2n}\}\leqslant E\{2^{\nu(n)}\}2^{-2n}=(e/4)^n,$$

by Stirling's formula

$$1/P\{\nu(n)=n\}=n!e^nn^{-n}=\sqrt{2\pi n}\left(1+\frac{1}{12n}+\frac{1}{288n^2}-\frac{139}{51840n^3}-\frac{571}{2488320n^4}+O\left(\frac{1}{n^5}\right)\right),$$

and by Laplace's method we get

$$\frac{P\{\nu(n)\leqslant n\}}{P\{\nu(n)=n\}}=\frac{e^nn!}{n^n}\int_n^{\infty}e^{-x}\frac{x^n}{n!}dx=\sqrt{\frac{n\pi}{2}}+\frac{2}{3}+\frac{\sqrt{2\pi}}{24n^{1/2}}-\frac{4}{135n}$$
$$+\frac{\sqrt{2\pi}}{576n^{3/2}}+\frac{8}{2835n^2}-\frac{139\sqrt{2\pi}}{103680n^{5/2}}+\frac{16}{8505n^3}-\frac{571\sqrt{2\pi}}{4976640n^{7/2}}+O\left(\frac{1}{n^4}\right).$$

(See Problem II-210 in G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1, Springer, Berlin, 1925, p. 80.)

By the formulas above,

$$b_n-a_n=\frac{2}{3}+\frac{8}{135n}-\frac{16}{2835n^2}-\frac{32}{8505n^3}+O\left(\frac{1}{n^4}\right)$$

as $n\to\infty$.

Also solved by D. W. Boyd (Canada), G. Gonnet (Canada), Attila Morte, L. E. Mattics, E. G. Straus, St. Olaf Problem Group, and M. M. Temme (Netherlands).

Disjoint Neighborhoods and Countable Local Bases

6274 [1979, 597]. *Proposed by F. S. Cater, Portland State University.*

Let S denote a topological space in which every compact set is closed, and let x and y be distinct points of S.

(1) Prove that x and y have disjoint neighborhoods if each of x and y has a countable local base.

(2) Show by example that x and y need not have disjoint neighborhoods if each element of S, other than x, has a countable local base.

By mistake, this problem also appeared as E 2806 [1979, 864]; see this issue, page 210, for the solution.

REVIEWS

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

We invite readers to submit reviews of significant recent college-level mathematics books. We especially encourage reviews based on classroom use and comparative reviews of several related books. Reviews should ordinarily not exceed two pages (per book) typed double spaced. Manuscripts of reviews as well as books submitted for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, MN 55057.

The Presidential Election Game. By Steven J. Brams. Yale University Press, New Haven, Connecticut, 1978. xix + 242 pp. $3.95 (Paper), $17.50. (Telegraphic Review, February 1979.)

Spatial Models of Election Competition. By Steven J. Brams. Birkhauser Boston, Inc., 380 Green Street, Cambridge, Massachusetts, 1979. v + 94 pp. $4 (Paper). (Telegraphic Review, August-September 1979.)

These books present some elementary mathematical models of the election process. They are written for "any conscientious reader with a reasonably good background in high school mathematics," according to Brams in the introduction to *The Presidential Election Game*. This description is a fair one. These books can be usefully employed in a course on mathematical modeling at an elementary level or as a supplementary source of ideas for a higher-level course where students might introduce more sophisticated techniques themselves.

The Presidential Election Game is not as cohesive a presentation as its title might lead one to believe. It is, rather, a collection of a number of different models for elections, woven together with the common theme of a presidential election. Each chapter focuses on one or two models applied to one aspect of a presidential election. With the exception of the last chapter, these topics follow the chronological order of the election process, including the demise of a president in Chapter 5.

The first chapter presents a model for the analysis of primaries—a spatial analysis of the positions candidates take with regard to the issues. The mathematical level of the book forces the consideration of only one-dimensional models. However, interesting consequences can be

developed at this level. For a course in mathematical modeling, higher dimensional analysis might provide some interesting problems.

Spatial Models of Election Competition—a monograph prepared for the Undergraduate Mathematics and Its Applications Project— is largely a somewhat expanded version of Chapter 1 of *The Presidential Election Game*. It also draws on a recent related paper [**3**]. For use in a modeling course it does have the advantage of having appropriate exercises.

Chapter 2 of *The Presidential Election Game* moves on to the conventions. Here a model, based on some elementary probability, is formulated to explain why and when delegates will shift their votes to one candidate or another—the "Bandwagon effect." Chapter 3 considers the general election, where the peculiar nature of the electoral college provides a rich structure for analysis. The chapter develops a measure of power for voters, related to but somewhat different from those found in [**1**] or [**5**]. It then uses the model to compare the electoral college system with popular election, concluding that voters in large states have a disproportionately large share of power under the former.

The fourth chapter analyzes coalition politics, in particular, party politics, in terms of some elementary ideas from game theory. Much of this chapter is inspired by Riker's work [**4**]. The fifth chapter provides a somewhat different angle. It uses the Supreme Court decision on Nixon's White House tapes as a vehicle for illustrating two-person nonzero-sum games. Although the connection of the model with reality is somewhat strained in this chapter, the analysis is instructive.

The final chapter of the book presents an alternative election method, one which the author has promoted for use in primary elections [**2**]: "approval voting." This is probably the most sophisticated chapter in the book and introduces many ideas about election methods. It is an interesting illustration of how precise mathematical techniques can be employed to analyze different election processes. The reader will have to decide whether the argument for approval voting is convincing.

The Presidential Election Game could be a useful adjunct to a course on mathematical modeling. It can illustrate how even mathematically simple models can lead to useful conclusions, and it should stir up many ideas in the mind of the curious reader. *Spatial Models*, though having less scope, can be used in the same way. I recommend their consideration for modeling courses, particularly at an elementary level.

References

1. John F. Banzhaf III, One man, 3.312 votes: A mathematical analysis of the electoral college, Villanova Law Review, 13 (1968) 304–332.

2. Steven J. Brams and Peter C. Fishburn, Approval voting, American Political Science Review, 72 (1978) 831–847.

3. Steven J. Brams and Philip D. Straffin, Jr., The entry problem in a political race, New York University, 1979.

4. William H. Riker, The Theory of Political Coalitions, Yale University Press, New Haven, Conn. 1962.

5. L. S. Shapley and Martin Shubik, A method for evaluating the distribution of power in a committee system, American Political Science Review, 48 (1954) 787–792.

CHRISTOPHER H. NEVISON, Colgate University

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, S(13), L??. Programmable Pocket Calculators. Henry Mullish, Stephen Kochan. Hayden, 1980, 254 pp, $8.95 (P). [ISBN: 0-8104-5175-1] Presents a number of programming techniques appropriate for programmable pocket calculators. Text focuses on the Hewlett-Packard family of calculators. AO

General, S*(13). A Book on Casino Craps, Other Dice Games & Gambling Systems. C. Ionescu-Tulcea. Van Nostrand Reinhold, 1981, ix + 149 pp, $8.95 (P); $12.95. [ISBN: 0-442-25725-2; 0-442-26713-4] First chapter gives a detailed description of how to play craps in a casino, indicating the house's edge and best bets, and includes an interesting section on cheating. Second chapter describes a variety of other dice games. Third and final chapter is a short discussion of probability, expectation and gambling systems. RSK

Precalculus, T(13: 1). Precalculus Mathematics. Thomas W. Hungerford, Richard Mercer, Sybil R. Barrier. Saunders Coll, 1980, xiv + 642 pp, $15.95. [ISBN: 0-03-020346-5] Standard precalculus topics in rational and transcendental functions for students with two or three years of high school mathematics. Extensive and detailed explanations and numerous examples in clear language show not only how, but why, the techniques work. Appears to be very accurate without being abstruse. Authors maintain that a typical student can more easily read and understand this longer discussion than a terse, compact text. Few real world applications are brought in and most of the problems seem to be routine drill. GHM

Precalculus, T(13: 1). Contemporary Analytic Geometry. Thomas L. Wade, Howard E. Taylor. Krieger, 1980, xiii + 324 pp, $17.50. [ISBN: 0-89874-034-7] A nice collection of topics in analytic and solid analytic geometry if a school has the luxury to offer such a course. Would like to see this used as a senior course in high school. (Original Edition, TR, October 1969.) LLK

Precalculus, T(13: 1). The Math Workshop: Elementary Functions. Deborah Hughes-Hallett. WW Norton, 1980, xiii + 562 pp, $15.95. [ISBN: 0-393-09033-7] Provides a brief review of algebra, then covers graphing, functions, logarithms and exponents. Introduces right-angle trigonometry before circular functions. LLK

Education, T(15: 1), S, P. How Children Learn Mathematics: Teaching Implications of Piaget's Research, Third Edition. Richard W. Copeland. Macmillan, 1979, vii + 419 pp, $13.95. [ISBN: 0-02-324780-0] Revision of 1974 edition (Second Edition, TR, December 1974). Four new chapters based on recent books by Piaget include memory and mathematics, knowing versus performing mathematics, chance and probability, and ordering and seriation. It remains a text for elementary teachers, emphasizing the child's stages of development and appropriate activities for helping children learn mathematics. PJ

History, P. L'Algèbre de la Logique, Deuxième édition. Louis Couturat. Albert Blanchard, 1980, 100 pp, (P). Exact reproduction of first edition published circa 1905. A leading French supporter of Russell's logicism, Couturat gives a thorough account of the "algebra of [propositional] logic" as invented by Boole and developed by Schröder up to the turn of the century. This republication contains absolutely no modern commentary and even lacks such basic information as the date of original publication. GHM

Foundations, P, L*. Wittgenstein on the Foundations of Mathematics. Crispin Wright. Harvard U Pr, 1980, xix + 481 pp, $30. [ISBN: 0-674-95385-1] A systematic account of Wittgenstein's later philosophy of mathematics (themes from Remarks on the Foundations of Mathematics) emphasizing the connections with his philosophy of language. Exposition is mainly sympathetic to Wittgenstein's views, arguing that his philosophy of mathematics is widely misconstrued when taken out of context of his general views (as in Philosophical Investigations). An index, analytical table of contents and page references to Wittgenstein's writings make this a very useful reference. GHM

Foundations, T(15–16: 1). Logic and Structure. Dirk van Dalen. Springer-Verlag, 1980, ix + 172 pp, $12.80 (P). [ISBN: 0-387-09893-3] First undergraduate course in mathematical logic for a general mathematical audience. Opts for a practical model-theoretic approach rather than a fastidious proof-theoretic one. Uses natural deduction to accord with intuition. Common mathematical structures are introduced as early as possible and emphasized throughout. Good coverage of main results of model theory. No coverage of recursion theory or incompleteness. Occasional awkward or ambiguous phraseology and nonstandard terminology. GHM

Foundations, P. Word Problems II, The Oxford Book. Ed: S.I. Adian, W.W. Boone, G. Higman. Stud. in Logic and Found. of Math., V. 95. North-Holland, 1980, x + 578 pp, $68.25. [ISBN: 0-444-85343-X] Twenty-four papers growing out of the conference on Decision Problems in Algebra held in Oxford in 1976. Two survey papers on algorithms used to solve decision problems and algorithmic problems for solvable groups. Research papers on embedding theorems, word problems for various classes of groups and algebras, cancellation theory, conjugacy problems, algebraically closed groups. List of seven open problems. KS

Foundations, P. The Kleene Symposium. Ed: Jon Barwise, H. Jerome Keisler, Kenneth Kunen. Stud. in Logic and Found. of Math., V. 101. North-Holland, 1980, xx + 425 pp, $61. [ISBN: 0-444-85-345-6] Proceedings of the symposium held June 18-24, 1978 at Madison, Wisconsin in honor of S.C. Kleene's 70th birthday. Research and expository papers on recursion theory and intuitionism. Includes brief biography of Kleene, plus summary and complete bibliography of his work. KS

Foundations, P. Science Without Numbers, A Defence of Nominalism. Hartry H. Field. Princeton U Pr, 1980, xiii + 130 pp, $16. [ISBN: 0-691-07260-4] Main thesis: using Hilbert's theory of geometry as a model, scientific theories can be nominalized, i.e., reaxiomatized so that no reference is made to mathematical entities. Consequently, to explain the application of mathematics to the physical world, "it is necessary to assume little more than that mathematics is consistent." KS

Combinatorics, P. Topics on Steiner Systems. Ed: C.C. Lindner, A. Rosa. Annals of Discrete Math., No. 7. North-Holland, 1980, ix + 349 pp, $68.25. [ISBN: 0-444-85484-3] An excellent collection of articles (both survey and research) on Steiner systems, including both algebraic and combinatorial aspects, which reflect the state-of-the-art and the direction in which research is moving. LCL

Combinatorics, T(13-14: 1). Digraphs: Theory and Techniques. D.F. Robinson, L.R. Foulds. Gordon, 1980, xv + 256 pp, $35. [ISBN: 0-677-05470-X] Introduction to directed graphs written at beginning college or senior high school level for students interested in applications to mathematics, engineering, business or the social or biological sciences. Assumes little background. Considerable attention given to algorithms for solving practical problems (e.g., branch and bound, critical path). Many exercises, with selected solutions. Chatty. GHM

Combinatorics, T(17: 1), S, P, L. Ramsey Theory. Ronald L. Graham, Bruce L. Rothschild, Joel H. Spencer. Wiley, 1980, ix + 174 pp, $21.95. [ISBN: 0-471-05997-8] A sophisticated and well-written introduction to Ramsey theory. The first four chapters are devoted to self-contained expositions of the central results of Ramsey theory. The fifth chapter deals with recent developments in the field and chapter six explores the influence of other areas of mathematics on Ramsey theory. Includes a sizable list of references. Does not include exercises. CEC

Combinatorics, P. Combinatorial Mathematics, Optimal Designs and Their Applications. Ed: J. Srivastava. Annals of Discrete Math., No. 6. North-Holland, 1980, viii + 391 pp, $68.25. [ISBN: 0-444-86048-7] A collection of 31 papers covering a wide range of topics in design theory and graph theory. Many of the papers are survey in nature. SG

Number Theory, P. Lecture Notes in Mathematics-800: Arithmétique des Algèbres de Quaternions. Marie-France Vignéras. Springer-Verlag, 1980, vii + 169 pp, $12.70 (P). [ISBN: 0-387-09983-2] Detailed examination of algebras of quaternions over local and global fields. Application to Riemann surfaces. Exercises and/or examples at the end of each chapter. JG

Number Theory, S. Von Fermat bis Minkowski: Eine Vorlesung über Zahlentheorie und ihre Entwicklung. W. Scharlau, H. Opolka. Springer-Verlag, 1980, xi + 224 pp, $18.90 (P). [ISBN: 0-387-10086-5] An account of the development of certain parts of number theory. Intended chiefly for prospective teachers, it contains a good deal of solid mathematics as well as the historical material. JD-B

Number Theory, P. Résultats Effectifs sur la Représentation des Entiers par des Formes Décomposables. Kálmán Győry. Pure and Appl. Math., No. 56. Queen's U, 1980, iii + 142 pp, $8.25 (P). A survey of recent work on effective solution of integral homogeneous polynomials; discusses work of Baker and Feldman on Thue's equation, Schmidt's work on norm forms, discriminant and index forms, and related topics. SG

Linear Algebra, T(14-15: 1). Lineare Algebra. Urs Stammbach. Teubner Stuttgart, 1980, 258 pp, (P). [ISBN: 3-519-00082-2] A demanding first text in linear algebra, intending to introduce the student not only to the material but also to the power and efficiency of abstract mathematical thinking. Most results are formulated abstractly in terms of (finite dimensional) vector spaces over an arbitrary field (without delving into the structure of fields per se). Most examples, of which there are a good number, are left to the reader to work out and seem to be incidental to the exposition. In German. GHM

Linear Algebra, T(15-16: 1). Generalized Inverses: Theory and Applications. Adi Ben-Israel, Thomas N.E. Greville. Krieger, 1980, xi + 395 pp, $24. [ISBN: 0-88275-991-4] The authors have limited their topics to generalized inverses of finite matrices, with brief extensions to infinite-dimensional spaces and to differential and integral operators in the last chapter. (Original Edition, TR, January 1975.) LLK

Linear Algebra, P. Non-Abelian Minimal Closed Ideals of Transitive Lie Algebras. Jack F. Conn. Princeton U Pr, 1981, 220 pp, $7.50 (P). [ISBN: 0-691-08251-0] A summary of the work of the author

and others on non-abelian minimal closed ideals of transitive Lie algebras over R and C. Contains necessary background material on transitive Lie algebras and derivations of transitive and simple algebras. SG

Algebra, P. Lecture Notes in Mathematics-772: Algebraic Structure of Knot Modules. Jerome P. Levine. Springer-Verlag, 1980, xi + 104 pp, $9.80 (P). [ISBN: 0-387-09739-2] Investigation of types of modules and product structures which arise as Alexander modules in the study of n-dimensional knots. Uses sequences of submodules and quotients, with emphasis on realizability theorems for such sequences. KS

Algebra, P. Ring Theory and Algebra III: Proceedings of the Third Oklahoma Conference. Ed: Bernard R. McDonald. Lect. Notes in Pure and Appl. Math., V. 55. Dekker, 1980, xvi + 422 pp, $49.75 (P). [ISBN: 0-8247-1158-0] Major papers by T.-Y. Lam, Richard Swan, O.T. O'Meara, David Eisenbud, and Irving Reiner, representing current trends in algebra and ring theory. LCL

Algebra, T(16-17: 2, 3), S, P, L. Algebra. Jacob K. Goldhaber, Gertrude Ehrlich. Krieger, 1980, xiv + 418 pp, $17.50. [ISBN: 0-88275-765-2] Intended as a text for first year graduate students, the book develops the basic theory for groups, rings, and fields with some more specialized and advanced results for each; including chapters on fields with real valuations and Noetherian and Dedekind Domains. The emphasis is on mappings, although category theory is not developed in a formal way. Includes exercises, bibliography, indices. (Original Edition, TR, April 1970; ER, August-September 1971 and April 1972.) JS

Algebra, P. Lecture Notes in Mathematics-825: Ring Theory, Antwerp 1980. Ed: F. van Oystaeyen. Springer-Verlag, 1980, vii + 209 pp, $14 (P). [ISBN: 0-387-10246-9] Proceedings of the "Second Week of Ring Theory at U.I.A." held at the University of Antwerp U.I.A., Antwerp, Belgium, May 6-9, 1980. JAS

Algebra, P. Cohomology of Completions. Saul Lubkin. Math. Stud., V. 42. North-Holland, 1980, xxx + 802 pp, $73.25 (P). [ISBN: 0-444-86042-8] A study of "the cohomology of the t-adic completion of cochain complexes of left A-modules," where A is a ring and t is an element of A. Brief bibliography, no index, substantial price for typescript. JAS

Algebra, P. An Introduction to Category Theory. V. Sankrithi Krishnan. North-Holland, 1981, x + 173 pp, $29.95. [ISBN: 0-444-00383-5] Definitely introductory, but also very definitely for mature readers; the approach is quite axiomatic. Each chapter provides a number of exercises and there is both a short index and a short bibliography. For the casual reader the introduction of a large number of special symbols with no index of the same is a serious flaw. JAS

Algebra, S(15-16), L. Group Tables. A.D. Thomas, G.V. Wood. Shiva Pub, 1980, (P). [ISBN: 0-906812-20-X] Provides multiplication table, character table, and subgroup lattice of each non-cyclic group of order at most 32. SG

Algebra, P. Lecture Notes in Mathematics-808: Ordres Maximaux au Sens de K. Asano. Guy Maury, Jacques Raynaud. Springer-Verlag, 1980, viii + 192 pp, $12.70 (P). [ISBN: 0-387-10016-4] An up-to-date account of the rings described in the title. SG

Calculus, T(13: 1). Mathematik heute: Grundkurs Analysis 2. Hermann Athen, Heinz Griesel. Hermann Schroedel, 1980, 113 pp, (P). [ISBN: 3-507-83083-3] Sequel to Grundkurs Analysis 1. Carefully planned introduction to basic integral calculus, up to simple applications and techniques of integration. Two chapters on differentiating transcendental functions are not applied to integration. Three-tone printing with text and exercises in parallel columns. Overall a very carefully reasoned didactic presentation. Suitable for high school (in German). GHM

Calculus, T(13: 2). A Short Calculus: An Applied Approach, Third Edition. Daniel Saltz. Goodyear, 1980, xii + 560 pp, $19.95. [ISBN: 0-87620-820-0] "The goal is to present some basic concepts and facts [no formal proofs] from calculus and to relate them whenever feasible to business, economics and the life sciences." Covers standard topics (including log and exp) up through partial derivatives and double integrals, with trigonometric functions reserved for the end. Emphasizes applications, though most of the applied problems do little more than state equations said to model a given situation and ask the student to solve. Still, the models are drawn from the literature and make for a realistic text. (TR, First Edition, June-July 1973; Second Edition, October 1977.) GHM

Real Analysis, T(15-16: 2), S, P*, L. Advanced Calculus. Harold M. Edwards. Krieger, 1980, xvii + 508 pp, $19.50. [ISBN: 0-89874-047-9] Unorthodox, but interesting and thorough in approach, using differential forms. Well-written, using motivational examples and drawings. Broad coverage of topics. Generous exercise sets, with full descriptions of solutions. No bibliography, but many references given in side-margin notes. (Original Edition, TR, October 1969.) MB

Real Analysis, S(15-16), P, L***.** The Generalized Riemann Integral. Robert M. McLeod. Carus Math. Mono., No. 20. MAA, 1980, xiii + 275 pp, $18. [ISBN: 0-88385-021-4] Beginning with the concept of the Riemann integral as developed in the usual calculus sequence, the ideas are extended so that the class of integrable functions include Lebesgue integrable functions and more. With a sense of building continuously on what they already know, readers come to the monotone and dominated convergence theorems and other results commonly reached only through the Lebesgue theory. All proofs are there, but beautiful organization aids the reader wanting to get the main ideas without

struggling with all the details. AWR

Real Analysis, T(16-17: 1, 2), S, L. Lebesgue Integration. Soo Bong Chae. Pure and Appl. Math., V. 58. Dekker, 1980, x + 314 pp, $35. [ISBN: 0-8247-6983-X] Well conceived, clearly written book. Nice historical and biographical notes including a paper by Lebesgue describing his work. Many exercises, an extensive bibliography; a grand text. Also, a bit expensive. PH

Complex Analysis, P. Lecture Notes in Mathematics-807: Fonctions de Plusieurs, Variables Complexes IV. Ed: François Norguet. Springer-Verlag, 1980, ix + 198 pp, $14.80 (P). [ISBN: 0-387-10015-6] Lectures from the Séminaire François Norguet, October 1977 to June 1979. Previous lectures (1970-1977) appear as Lecture Notes 409, 482, and 670. JAS

Complex Analysis, S(18). Vorlesungen über Riemannsche Flächen. Kurt Strebel. Vandenhoeck & Ruprecht, 1980, 120 pp, DM 18,80 (P). [ISBN: 3-525-40145-0] A proof of the fundamental theorem on conformal representation of simply-connected Riemann surfaces, using Dirichlet's principle and not integration on Riemann surfaces. JD-B

Differential Equations, S(17-18). Lectures on Wave Propagation. G.B. Whitham. Springer-Verlag, 1979, v + 148 pp, $8 (P). [ISBN: 0-387-08945-4] The theory of water waves, emphasizing results concerning waves on a sloping beach. Includes a basic introduction to the theory of characteristics and shock waves. AO

Differential Equations, T(17-18: 1, 2), S, L. Ordinary Differential Equations. Jack K. Hale. Pure and Appl. Math., V. 21. Krieger, 1980, xvi + 361 pp, $27.50. [ISBN: 0-89874-011-8] Misprints have been eliminated, presentations of some material clarified, section on integral manifolds enlarged, new material on Hopf bifurcation and related topics introduced, appendix on almost periodic functions rewritten (using the modern definition of Bochner), and references on recent developments added. (Original Edition, TR, April 1970; ER, December 1971.) JK

Differential Equations, P. Lecture Notes in Mathematics-791: Differential Operators for Partial Differential Equations and Function Theoretic Applications. Karl Wilhelm Bauer, Stephan Ruscheweyh. Springer-Verlag, 1980, v + 258 pp, $16.80 (P). [ISBN: 0-387-09975-1] This volume contains two papers. The first is a survey of the known results concerning the representation of solutions of partial differential equations by differential operators. The second is on the function theory of the Bauer-Peschl equation. AO

Numerical Analysis, T(17-18: 1), P, L. Approximation Theory and Numerical Methods. G.A. Watson. Wiley, 1980, x + 229 pp, $24.95. [ISBN: 0-471-27706-1] After an initial chapter that sets approximation in the context of finding points of a set nearest to a fixed element of a normed linear space, the rest of the book is set in the language of classical function theory rather than functional analysis. Most of the book concentrates on standard linear problems with L_p norms. Should be attractive as a text (prerequisites of real analysis, linear algebra, some linear programming) or as a reference for undergraduate honors work. AWR

Numerical Analysis, P. Numerical Methods. Ed: P. Rózsa. North-Holland, 1980, 631 pp, $87.75. [ISBN: 0-444-85407-X] Proceedings of the colloquium on numerical methods held in Keszthely, Hungary, September 5-10, 1977. Emphasis is on numerical algebra and numerical solutions to partial differential equations. JAS

Functional Analysis, P. Lecture Notes in Mathematics-786: Infinite Matrices of Operators. Ivor J. Maddox. Springer-Verlag, 1980, v + 122 pp, $9.80 (P). [ISBN: 0-387-09764-3] A survey of the main developments which have occurred since 1950. AO

Functional Analysis, T(18), P. Operator Ideals. Albrecht Pietsch. Math. Lib., V. 20. North-Holland, 1980, 451 pp, $73.25. [ISBN: 0-444-85293-X] Contains both the abstract theory (Parts I and II) and the more specialized theory related to particular examples (e.g., Hilbert space) or ideals having interesting properties and applications (Parts III and IV). Part V is devoted to applications in different branches of functional analysis and probability theory. Large 500 item bibliography. LCL

Optimization, T*(16-17: 1), L.** Practical Methods of Optimization, Volume 1: Unconstrained Optimization. R. Fletcher. Wiley, 1980, viii + 120 pp, $24.50. [ISBN: 0-471-27711-8] An attractive, introductory text by an author with much practical experience. Emphasis is on understanding the considerations involved in choosing practical algorithms to handle some standard problems; some comparative numerical studies are given. Author's experience is drawn upon to present questions taking students to the heart of the subject. Volume II will deal with constrained optimization. AWR

Optimization, S(15-16), L. Combinatorial Methods of Discrete Programming. László Béla Kovács. Akademiai Kiado, 1980, 283 pp, $25. [ISBN: 963-05-2004-4] A clearly written introduction which unfortunately lacks exercises; discusses the standard problems from several points of view; concludes with a survey of recent directions in discrete programming. SG

Analysis, T(17: 1, 2), P. An Introduction to Variational Inequalities and Their Applications. David Kinderlehrer, Guido Stampacchia. Pure and Appl. Math., V. 88. Acad Pr, 1980, xiv + 313 pp, $35. [ISBN: 0-12-407350-6] Graduate level text or reference for mathematically literate engineers, economists, etc. One chapter devoted to applications, historical notes and comments, extensive

bibliography. Problems would need supplementing if used as a text. PH

Analysis, T(18), S, P. Lecture Notes in Mathematics-801: Weakly Compact Sets. Klaus Floret. Springer-Verlag, 1980, vii + 123 pp, $9.80 (P). [ISBN: 0-387-09991-3] This book has three foci: the theorems on countable compactness, on sequential compactness, and the supremum of linear functionals. The linking element is A. Grothendieck's interchangeable double-limit property. It is clearly written, with abundant exercises. MU

Analysis, P. Quantitative Approximation. Ed: Ronald A. DeVore, Karl Scherer. Acad Pr, 1980, xi + 324 pp, $22. [ISBN: 0-12-213650-0] A collection of 26 papers presented to a symposium on quantitative approximation held in Bonn, West Germany in August, 1979. Reproduced from typewriter copy, the articles are state-of-the-art, suggesting many open problems for research. AWR

Analysis, T(16-17: 1, 2), S. Measure-Theoretic Probability. Henry A. Krieger. U Pr of America, 1980, xi + 382 pp, $12.50 (P); $20.50. [ISBN: 0-8191-1228-3; 0-8191-1229-1] An introduction to measure and integration motivated by probability in Chapters 1-5. Chapters 6-12 consider applications to probability. Exercises and examples, adequate. Print quality, poor. Limited bibliography. PH

Algebraic Geometry, P. Lecture Notes in Mathematics-815: Simple Singularities and Simple Algebraic Groups. Peter Slodowy. Springer-Verlag, 1980, xi + 175 pp, $11.80 (P). [ISBN: 0-387-10026-1] Proves results on the connections between rational double points and simple Lie algebras within the framework of algebraic geometry over algebraically closed fields. Revised and translated version of Einfache Singularitäten und einfache algebraische Gruppen. KS

Algebraic Geometry, S(16-17), L. Studies in Algebraic Geometry. Ed: A. Seidenberg. Stud. in Math., V. 20. MAA, 1980, xi + 143 pp, $16. [ISBN: 0-88385-120-2] A collection of four essays whose aim is to provide a quick peak at this vast subject. The essays and authors are: "Initial results in the theory of linear algebraic groups" by M. Rosenlicht; "The connectedness theorem and the concept of multiplicity" by B.L. van der Waerden; "Space curves as ideal-theoretic complete intersections" by J. Ohm; and "Charles's enumerative theory of conics: a historical introduction" by S. Kleiman. SG

Algebraic Geometry, P. Lecture Notes in Mathematics-817: Schottky Groups and Mumford Curves. Lothar Gerritzen, Marius van der Put. Springer-Verlag, 1980, viii + 317 pp, $19.50 (P). [ISBN: 0-387-10229-9] An exposition of work of the authors D. Goss and F. Herrlich on Schottky groups and the corresponding Mumford curves. SG

Algebraic Geometry, P. Lecture Notes in Mathematics-820: The Real Analytic Theory of Teichmüller Space. William Abikoff. Springer-Verlag, 1980, vii + 144 pp, $11.80 (P). [ISBN: 0-387-10237-X] From the introduction: "These notes are based on a series of four lectures delivered by Lipman Bers and the author in the "Seminaire sur les diffeomorphismes des surfaces d'apres Thurston." ...Almost all...is a background to or a modest exposition of Bers' work on the problem of moduli." SES

Geometry, S(13-14), L. Plane Figures and Sections: How to Construct Them Given Specific Conditions. P.V. Pritulenko. Trans: Vladimir Shokurov. MIR Pub, 1980, 164 pp, $7. Four chapters on new constructions in descriptive geometry. Should be of interest to workers in the development, analysis and synthesis of spatial mechanisms. For example, Chapter 2 covers the construction of views of plane figures given their true size and the plane view of a similar figure, with the required figure in the same plane as the given figure, or in a plane parallel to it. Some problems. No solutions. Nine references, all in Russian. Over one hundred striking drawings. JK

Geometry, S(16), P.** Problems in Discrete Geometry, 1980, Fifth Edition. William Moser (Dept. of Math., McGill U., 805 Sherbrooke St. W., Montreal, Quebec H3A 2K6), 1980 (P). At first 14, then 20, then 22, then 28 and now 34 challenging open-ended problems in discrete geometry. In this edition, each problem has been assigned a monitor to whom any information, question or comment should be sent. A welcome new feature is the editor's offer to provide copying service ($.10 per page) for many of those bibliographic entries which are not otherwise accessible to readers. (Fourth Edition, TR, March 1980.) JK

Geometry, S(18), P. Collinearity-preserving Functions between Desarguesian Planes. David S. Carter, Andrew Vogt. Memoirs No. 235. AMS, 1980, v + 98 pp, $5.20 (P). [ISBN: 0-8218-2235-7] A characterization of all collinearity-preserving functions from one affine or projective Desarguesian plane into another is obtained using concepts from valuation theory. These results permit one or both planes to be affine and include cases where the range contains a triangle but no quadrangle. CEC

Geometry, P. Lecture Notes in Mathematics-792: Geometry and Differential Geometry. Ed: R. Artzy, I. Vaisman. Springer-Verlag, 1980, vi + 443 pp, $24.50 (P). [ISBN: 0-387-09976-X] Proceedings of the conference held at the University of Haifa, Israel, March 18-23, 1979. JAS

Geometry, T(17-18). Konvexe Mengen. K. Leichtweiss. Springer-Verlag, 1980, 330 pp, $31.90 (P). [ISBN: 0-387-09071-1] A sophisticated introduction to the theory of convex sets in n-dimensional real affine space. Exercises and a small bibliography. JD-B

Topology, S(15-17), L. Topologie. Klaus Jänich. Springer-Verlag, 1980, ix + 215 pp, $13 (P). [ISBN: 0-387-10183-7] Overview of set theoretic topology for the nonspecialist. Relaxed,

interesting style. PH

Probability, S. Search Games. Shmuel Gal. Math. in Sci. and Eng., V. 149. Acad Pr, 1980, xiv + 216 pp, $20. [ISBN: 0-12-273850-0] Written for people interested in search and minimax problems, this book calls upon basic game theory and probability as tools for finding optimal search trajectories to locate a target. Printed from photo of typed manuscript. AWR

Probability, S(17-18), P. Stochastic Integration. Michel Metivier, J. Pellaumail. Prob. and Math. Stat. Acad Pr, 1980, xii + 196 pp, $25. [ISBN: 0-12-491450-0] A self-contained exposition of the theory of stochastic integration requiring only a basic knowledge of measure-theoretic probability theory. Includes a chapter on infinite-dimensional stochastic integration and an extensive bibliography. AO

Statistics, P. Engineering Applications of Correlation and Spectral Analysis. Julius S. Bendat, Allan G. Piersol. Wiley, 1980, xiv + 302 pp, $29.95. [ISBN: 0-471-05887-4] Concerned with general interpretations and applications of random data analysis with emphasis on the use of correlation and spectral density functions. Supplements the authors' 1971 book Random Data: Analysis and Measurement Procedures (TR, August-September 1974). RSK

Statistics, T(13: 1). An Introduction to Data Analysis. Bruce D. Bowen, Herbert F. Weisberg. Freeman, 1980, xi + 213 pp, $15.95; $7.95 (P). [ISBN: 0-7167-1173-7; 0-7167-1174-5] Discussion of the most common statistical techniques, including both descriptive and inferential statistics. Written to explain general principles; very limited exposure to problem solving. LCL

Statistics, P. Cluster Analysis Algorithms for Data Reduction and Classification of Objects. Helmuth Späth. Trans: Ursula Bull. Halsted Pr, 1980, 226 pp, $56.95. [ISBN: 0-470-26946-4] In the Ellis Horwood Series in Computers and Their Applications. Translation and revision of the author's 1977 second German edition. Explains and gives Fortran programs for the most important algorithms of cluster analysis, with worked examples illustrating their use. Good bibliography. RSK

Statistics, T*(16-17; 1), S, P*, L. Applied Linear Regression. Sanford Weisberg. Wiley, 1980, xii + 283 pp, $24.95. [ISBN: 0-471-04419-9] In the Wiley Series in Probability and Mathematical Statistics. Intended for computer-using practitioners, or as a text for a second or third course in statistics. Well-written but concise presentation of many of the latest techniques in regression analysis, illustrated with real data. Central themes are model building, assessing fit and reliability, and drawing conclusions. Good set of references. RSK

Statistics, T(17: 1). Life Testing and Reliability Estimation. S.K. Sinha, B.K. Kale. Halsted Pr, 1980, vii + 196 pp, $16.95. [ISBN: 0-470-26911-1] Presumes two years of calculus and a course in mathematical statistics. Concerned with the theory and methodology of inference, mainly estimation (including Bayesian) and tests of hypotheses, in the area of life testing. RSK

Statistics, T(16: 1, 2). Survival Models and Data Analysis. Regina C. Elandt-Johnson, Norman L. Johnson. Wiley, 1980, xvi + 457 pp, $34.95. [ISBN: 0-471-03174-7] In the Wiley Series in Probability and Mathematical Statistics. Deals primarily with the treatment of mortality data. Divided into four parts: survival measurements and concepts, mortality experiences and life tables, multiple types of failures, and some more advanced topics. Good sets of chapter references. RSK

Statistics, P. Testing the Constancy of Regression Models over Time. Peter Hackl. Appl. Stat. and Econometrics, V. 16. Vandenhoeck & Ruprecht, 1980, 132 pp, DM 36 (P). [ISBN: 3-525-11247-5] Restricted to models applied to observations on continuous variables made at equally distant time points. Primarily concerned with moving sum techniques, and their advantages over cumulative sum techniques. Good bibliography. RSK

Statistics, T(15-17:1, 2), P, L. Practical Nonparametric Statistics, Second Edition. W.J. Conover. Wiley, 1980, xiv + 493 pp, $25.95. [ISBN: 0-471-02867-3] In the Wiley Series in Probability and Mathematical Statistics. Revision of the author's successful 1971 First Edition (TR, May 1973). Material has been updated, problems have been reorganized into theoretical and applied sections, review problems have been added, but the essential format and substance of the book are unchanged. RSK

Statistics, S(17-18), P*. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. David A. Belsley, Edwin Kuh, Roy E. Welsch. Wiley, 1980, xv + 292 pp, $21.95. [ISBN: 0-471-05856-4] In the Wiley Series in Probability and Mathematical Statistics. Detailed treatment emphasizing diagnosis of potential data problems rather than inference or curve fitting. Gives both theoretical bases for the diagnostic techniques and means of implementation. Good bibliography. RSK

Statistics, T*(13-14: 1, 2). Statistics, A First Course, Third Edition. John E. Freund. P-H, 1981, xiv + 466 pp, $17.95. [ISBN: 0-13-845958-4] An attractive presentation of the usual topics. Presupposes no college level mathematics. Many changes from the Second Edition. (First Edition, TR, August-September 1970; Second Edition, TR, May 1976.) FLW

Statistics, P*. Developments in Statistics, Volume 3. Ed: Paruchuri R. Krishnaiah. Acad Pr, 1980, xiv + 254 pp, $35. [ISBN: 0-12-426603-7] Four invited papers giving "authoritative reviews of the

present state-of-the-art on some aspects of asymptotic expansions [in parametric statistical theory], [orthogonal] models for contingency tables, statistical concepts in economic analysis, and path analysis." Good sets of references. RSK

Computer Science, S(15-16), P. Software Configuration Management. Edward H. Bersoff, Vilas D. Henderson, Stanley G. Siegel. P-H, 1980, xiv + 385 pp, $24.95. [ISBN: 0-13-821769-6] Presents a set of principles and techniques that can be used as a guide in the procurement and/or development of large-scale software systems. AO

Computer Science, T(17-18: 1), S, P. Lecture Notes in Computer Science-92: A Calculus of Communicating Systems. Robin Milner. Springer-Verlag, 1980, vi + 171 pp, $11.80 (P). [ISBN: 0-387-10235-3] CCS, a formal calculus, describes concurrent systems and provides a proof methodology for such systems. In this volume the expressive capabilities of CCS are demonstrated by a series of case studies with exercises and extensive references. JL

Computer Science, S*(15), P, L*. Error-Free Computation: Why It Is Needed and Methods For Doing It. Robert Todd Gregory. Krieger, 1980, v + 152 pp, $7.50 (P). [ISBN: 0-89874-240-4] An expository monograph which deals with numerical mathematics by looking at ill-conditioned problems, numerically unstable algorithms, the scaling problem for linear systems, and the uses of residue and p-adic arithmetic for obtaining exact computational results. Easy to read. Hard proofs are omitted, but referenced. CEC

Computer Science, S(13), L. Computers & Social Controversy. Tom Logsdon. Computer Sci Pr, 1980, xii + 397 pp, $17.95 [ISBN: 0-914894-14-5]; Workbook. vii + 123 pp, (P). A well-documented book on many current issues. Each chapter has a bibliography and some chapters also include a suggested reading list. LLK

Computer Science, T(14: 1), S. The Nature of Computation: An Introduction to Computer Science. Ira Pohl, Alan Shaw. Computer Sci Pr, 1981, xii + 397 pp, $16.95. [ISBN: 0-914894-12-9] The history and uses of computers along with social implications and controversies. Also provides an initial literacy in the language and methods of computer science. LLK

Applications (Biology), T*(18: 1), S*. Deterministic Mathematical Models in Population Ecology. H.I. Freedman. Pure and Appl. Math., V. 57. Dekker, 1980, x + 254 pp, $29.75. [ISBN: 0-8247-6653-9] A good introduction to the dynamics aspect of mathematical ecology from a mathematician's point of view. Most models included are given by autonomous ordinary differential equations or difference equations. Predator-prey, competition, and cooperation models are considered. Exercises include open problems. Extensive bibliography. SES

Applications (Engineering), S(17-18), P. The Excitation and Propagation of Elastic Waves. J.A. Hudson. Cambridge U Pr, 1980, viii + 226 pp, $32.50. [ISBN: 0-521-22777-1] An exposition of the fundamental results of the linearized theory of elastodynamics from the point of view of the propagation of transient pulses in an isotropic material which may be regarded as unbounded. AO

Applications (Engineering), T(16-17: 1). Computer Logic, Testing and Verification. J. Paul Roth. Computer Sci Pr, 1980, xx + 176 pp, $23.95. [ISBN: 0-914894-62-5] An exposition of the cubical calculus and its use in logic design, minimization, testing, embedding and verification. Each chapter includes problems and a bibliography. AO

Applications (Engineering). Information Linkage Between Applied Mathematics and Industry II. Ed: Arthur L. Schoenstadt, et al. Acad Pr, 1980, xii + 293 pp, $20. [ISBN: 0-12-628750-3] Proceedings of the second in a series of symposia held at the Naval Postgraduate School in an attempt to bring applied mathematicians and engineers together for a working session to exchange ideas and experiences. This workshop focused on linear systems, and occurred shortly after the release of the LINPACK library of linear systems software. Contains texts of five invited addresses, plus abstracts on texts of numerous contributed papers. LAS

Applications (Physics), P. Lecture Notes in Physics-129: Geometrical and Topological Methods in Gauge Theories. Ed: J.P. Harnad, S. Shnider. Springer-Verlag, 1980, viii + 154 pp, $14 (P). [ISBN: 0-387-10010-5] Contains the texts of eleven papers presented at the Canadian Mathematical Society Summer Research Institute which took place at McGill University from September 3-8, 1979. AO

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK T. KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

Assistant Professor *Glenn Allinger* of Montana State University has been promoted to Associate Professor.

Associate Professor *Brian Alspach* of Simon Fraser University has been promoted to Professor.

Walter Beck, formerly of Wartburg College, has been appointed Assistant Professor of Mathematics at the University of Northern Iowa.

Professor *Dorothy Carpenter* of Ashland College (Ohio) retired in June 1980.

Jeffrey Dinitz, a recent graduate of Ohio State University, has been appointed Assistant Professor at the University of Vermont.

Milton P. Eisner, formerly at J. Sargeant Reynolds Community College, has been named Assistant Professor at Mount Vernon College.

Professor *Bernard Greenspan* of Drew University has retired with the rank of Professor Emeritus.

Mrs. *Patricia Hoover* has retired from the faculty of Duquesne University.

Dr. *V. Linis* of the University of Ottawa was given a 1980 Teaching Award of the Ontario Confederation of University Faculty Associations.

Assistant Professor *John I. Moore*, Jr., of the Citadel has been promoted to Associate Professor.

Associate Professor *Barnet M. Weinstock* of the University of North Carolina at Charlotte has been promoted to Professor and is now acting chairman of the Department.

Albert Yu-Ming Chi, formerly of Emporia State University, has been named Assistant Professor of Mathematics at Lebanon Valley College.

University of Akron: Assistant Professor *Wilbur P. Veith* has been promoted to Associate Professor. *George L. Szoke* has completed the Ph.D. in Engineering at the Technical University of Budapest.

California State Polytechnic University, Pomona: *Carol Smith*, a graduate of the University of Texas specializing in Mathematics Education, has been appointed Assistant Professor. Associate Professors *Hasan Çelik* and *John R. Fisher* have been promoted to Professor.

University of Delaware: Newly appointed Assistant Professors include *Dennis F. Karney*, formerly of the University of Illinois, and *Michael E. Gage*, formerly of the University of Washington. Assistant Professor *Ronald D. Baker* has been promoted to Associate Professor.

Gordon College: *Richard Stout*, formerly of Grove City College, has been appointed Assistant Professor. Professor *Harold Heie* has resigned to become Vice-President for Academic Affairs at Northwestern College, Orange City, Iowa. Assistant Professor *Walter Stangl* has resigned to pursue studies at the Conservative Baptist Theological Seminary.

Jarvis Christian College: *William P. Riemen*, formerly Assistant Professor at the University of Guam, has been appointed Associate Professor. Associate Professor *Vaclav Konecny* has resigned to take a position at Ferris State University, Michigan.

Kearney State College: *Charles G. Pickens* is the new chairman of the Department of Mathematics, Statistics, and Computer Science. He has been a member of the department since 1960. Associate Professor *Richard L. Barlow* has been promoted to Professor of Statistics.

Lewis and Clark College: Associate Professor *Roger B. Nelsen* has been named Department Chairman. Professor *Elvy L. Fredrickson* retired on June 15, 1980, after 32 years of service with the title Professor Emeritus.

McNeese State University: Professor *Harlin W. Brewer* has been named Head of the Department of Mathematical Sciences, succeeding *Patrick L. Ford*, who has retired.

University of Missouri, St. Louis: Recent appointments include Assistant Professor *Ronald M. Dotzel*, formerly of the University of Texas, and Associate Professor *Marjory Johnson*, formerly of the University of South Carolina.

University of Michigan: Professor *Cecil J. Nesbitt* has retired with the rank of Professor Emeritus. Professor Emeritus *Raymond L. Wilder* received an honorary doctorate from the University of Michigan in May 1980.

University of Missouri, Rolla: Associate Professor *John C. Kieffer* has been promoted to Professor. He is now an Associate Editor of IEEE Transactions on Information Theory.

North Carolina State University: *J. Rodriguez*, formerly of the University of Maryland, has been appointed Assistant Professor. Associate Professor *J.B. Wilson* has been promoted to Professor.

Northern Michigan University: Assistant Professor *Lawrence N. Meyerson* has resigned to pursue the study of law at Seton Hall, N.J. Professor *Jane O. Swafford* has been named Dean of Graduate Studies. Professor *Theodore A. Eisenberg* is on academic leave at Ben Gurion University of the Negev, Beersheva, Israel.

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

FALL MEETING OF THE NORTH CENTRAL SECTION

The North Central Section of MAA met October 24 and 25, 1980, at North Dakota State University in Fargo. Prof. *Robert Ellis* of the University of Minnesota gave the first invited address, "Topological Dynamics: What it is and Where it is Going." Saturday morning talks included "Inverse of the Variance-Covariance Matrix for a Special Case" by *Dave Smith* of North Dakota State University; "A Mathematical Model of a Population Problem with Migration" by *Roger Avelsgaard,* Bemidji State University; "Graph Theory and Some Elementary Linear Algebraic Proofs" by *Michael Doob* of the University of Manitoba; "A Property That Very Few Triangles Have" by *Walter Fleming* of Hamline University; "Mathematics in the Technical Institutes and Community Colleges in New Zealand" by *H. Offenberger* of Anoka Ramsey Community College/Wellington Polytechnic, New Zealand.

A second invited address was given by Prof. *Donald G. Saari,* Northwestern University. His topic was "Computation of Fixed Points and Singularity Theory." Other speakers were *Joe Konhauser,* Macalester College, "On U.S. Patent No. 4,074,778--the Hole Story"; *Martha Schue,* Carleton College student, "Modeling a Waste Treatment Plant, or, Garbage In, Garbage Out"; *Douglas Dunham,* University of Minnesota, Duluth, "Repeating Patterns of the Hyperbolic Plane"; *Keith Pierce,* University of Minnesota, Duluth, "The Computer as Algebraist"; and *James Olsen,* North Dakota State University, "Ergodic Theorems for Subadditive Processes."

The meeting closed with "Little Harmonic Labyrinth," a theatrical adaptation of a dialogue from the book GODEL, ESCHER, BACH. It was presented by *Paul Fjelstad* and students from St. Olaf College.

FALL MEETING OF THE NEW JERSEY SECTION

The fall meeting of the New Jersey Section of the MAA was held on Saturday, October 25, 1980, at Union College in Cranford, New Jersey. Forty-seven registrants attended the meeting. *Jean Lane,* the section chairperson, presided.

The first speaker was *Edward Fisher* of Bell Labs in Murray Hill, N.J., whose talk was entitled "Future Directions for Logic and Algebra in Computer Science." This was followed at 11:30 by the business meeting where the new section officers were nominated and elected. The new section officers are: Chair-elect, *Susan G. Marchand* (Kean College); Vice-Chair for Speakers, Sister *Stephanie Sloyan* (Georgian Court College); Treasurer, *Judith B. Seery* (Bell Labs, Murray Hill). Officers re-elected to their present position were: Vice-Chair for Innovations, *Dorothy Wolff* (Caldwell College) and Vice-Chair for Two-Year Colleges, *Margaret Piedem* (Somerset County College). The last talk of the day, "Self-Reproducing Programs," was then presented by *Donald Kreider* of Dartmouth College.

NOVEMBER MEETING OF THE SEAWAY SECTION

The Seaway Section of MAA met November 7 and 8, 1980, at Daemen College, Amherst, N.Y. The followsing talks were presented:

"Sharks in the Mediterranean during WWI," *Arthur Deacon,* Syracuse University

"Actuarial Sciences in the College Mathematics Department," *Bette Warren,* SUNY, Binghamton

"Generalized Bounded Variation," *Michael Schramm,* Syracuse University

"Linear Algebra Made Difficult," *Paul Halmos,* Indiana University

"Constructive Trigonometry," *Shirley Stanlen,* Schenectady CCC and *Eugene Mozier,* La Salle Institute.

"A Mean Value Theorem in Higher Dimensions," *Richard O'Neil,* SUNY Center, Albany

"A Psychology of Mathematical Problem Solving," *Stephen West,* SUNY, Geneseo

"The Mathematics of Personnel Allocation: The Vices of Linearity," *John P. Mayberry,* Brock University

"Casino Gambling," *John Slivka,* Buffalo State College

"Recent Developments on Summability Invariants," *S.C. Chang,* Brock University

ANNUAL MEETING OF THE NORTHEASTERN SECTION

The twenty-sixth annual meeting of the Northeastern Section of MAA was held at Merrimack College Andover, Massachusetts on November 22, 1980. In the morning session Prof. *Richard A. Rasala* of Northeastern University spoke on "Esher's Periodic Drawings," and Prof. *Gian-Carlo Rota* of MIT gave the Christie Lecture, "The Fall and Rise of Invariant Theory."

At the business meeting officers were elected as follows: Vice-chairman: *James E. Ward,* Bowdoin College; Secretary-Treasurer: *Shirley A. Blackett,* Northeastern University; Two-Year College Representative: *Nancy Myers,* Bunker Hill Community College. The Section expressed its sincere thanks to *George Best* for his many years of service as Secretary-Treasurer of the Northeastern Section. Future meetings and short courses were announced.

The afternoon program included a panel discussion on "Student Reality, Faculty Abstraction and Intellectual Development, or Reality, Abstraction and Modeling in the Freshman Year." Participants were *Steven K. Ingram,* Norwich University; *Michael Olinick,* Middlebury College; and *Alan Natapoff,* MIT. The program ended with simultaneous talks, "Re-entry Mathematics and the Non-Traditional Student," by *Jean B. Smith* of Middlesex Community College, Connecticut; and "Mathematical and Statistical Applications in Forestry," by *Homer T. Hayslett, Jr.,* of Colby College.

Report of the Committee on Improving Remediation Efforts in the Colleges
Mathematical Association of America

July, 1980
(Approved by the Board of Governors on August 20, 1980)

It is well documented that large numbers of freshmen entering the colleges and universities of this country lack adequate preparation in mathematics. The 1975-76 Report of the Survey Committee of the Conference Board of the Mathematical Sciences showed 141,000 remedial enrollments at four-year institutions and 245,000 at two-year institutions in 1975. An informal survey of 20 large two-year public colleges and 25 large state-supported universities, made in 1979 by the Committee on Improving Remediation Efforts in the Colleges, showed that almost all institutions are providing beginning courses that include elementary algebra and often arithmetic. This response suggests that the 1980-81 Survey will indicate sharp increases in remedial enrollments since 1975, at least at the four-year institutions. Indeed, the February 1980 Notices reports a 29% increase in course enrollments below calculus between fall 1978 and fall 1979 in the top 27 ACE ranked mathematics departments (Rung, Donald C., "Changes in Enrollments, Class Sizes, and Size of Faculty," Notices 27 (February, 1980), 176-177).

The Committee on Improving Remediation Efforts in the Colleges was formed in Autumn 1978 by the Executive Committee of the MAA in response to the 1978 PRIME 80 recommendations. The committee has four charges from the PRIME 80 recommendations:

1. to analyze the content of courses now being given for remediation purposes, ascertain from the faculty teaching these courses the areas of difficulty, and prepare recommendations that would assist these teachers in helping students overcome their deficiencies

2. to address the problem of development of appropriate and motivating curricular materials, especially source materials on applications

3. to involve extensively two-year college faculty and other faculty teaching courses of this kind in the study and in the development of materials

4. to urge the Sections of the MAA to provide workshops at their sectional meetings on the problems arising from insufficient preparation of students.

First, a word about a word: the Committee has found that the word "remedial" is emotionally charged for many people and some colleagues would prefer that it not be used to describe instruction for underprepared students in mathematics. Furthermore, courses developed for these students have different (and sometimes almost nonintersecting) contents at different institutions. Nonetheless, the Committee has not found a less offensive phrase and in this report will use the word "remedial" to refer to (a) those freshmen students who are not prepared to begin the mathematics courses required in their degree programs and (b) mathematics courses designed to prepare those students for entrance into the regular required courses in degree programs.

Before writing this report, the Committee spent several months studying current programs at junior colleges, four-year colleges, and universities. We found no institution that has not been affected by the lack of preparation of large numbers of students. There is predictably a great deal of diversity among the programs that have been developed, yet the Committee feels that it can identify some characteristics of successful programs; these will be listed in a later section.

For many years, the two-year institutions carried much of the burden of remedial instruction in mathematics. However, four or five years ago many four-year colleges and universities began to develop or expand remedial programs and, indeed, many other schools are now initiating new programs in response to at least two new pressures. First, many schools anticipate more vigorous recruitment of students and want to be sure that their curricula provide access to degree programs for all potential students. Second, some universities and many arts and sciences colleges have recently established or soon will establish a mathematics requirement for graduation, and mathematics departments are being charged to develop courses that will prepare students to meet this requirement. Thus, it appears that even more attention will be given to these matters by many departments in the next few years.

The Committee has found many institutions where faculty feel the remediation efforts are doing little good and, indeed, their data confirm this feeling. Teachers are frustrated by the low-level instruction and departments are hard pressed to absorb the costs of increased loads. On the other hand, in some schools the Committee has found considerable optimism. Individual faculty members report that they can realistically set higher goals for this audience than they first thought, that they have learned a great deal about the teaching of elementary mathematics from their involvement with remedial students, and that the development of remedial courses has sometimes resulted in the modification and improvement of other undergraduate courses. These individuals welcome the opportunity to teach good mathematics (albeit elementary) to an audience who might otherwise have taken no mathematics in their college programs.

The Committee has found programs which teach the technical skills needed in later courses and at the same time provide for conceptual understanding and for analytic reasoning; data indicate that large numbers of students are moving successfully from these programs into the mathematics courses required in degree programs. Some schools are using the instruction of remedial students as part of their program for the

training of teachers and are convinced that it is very important to involve future teachers with students who failed to learn pre-college mathematics.

In spite of the positive aspects of remedial efforts in some departments and the likelihood that the pressures to provide this instruction will not diminish for some time, the Committee feels strongly that the proper place for instruction at this level is in the secondary schools. The learning of elementary mathematics is best done over a period of several years. No highly concentrated program contained in a few months can substitute for the opportunity to deal with the ideas of elementary mathematics over an extended time. Several programs have been developed to encourage more adequate preparation in mathematics at the pre-college level. The National Council of Teachers of Mathematics and the MAA have made strong recommendations on college preparatory mathematics in the publication "Recommendations for the Preparation of High School Students for College Mathematics Courses." We shall cite other examples which we think are having significant effects and which indicate appropriate professional effort in solving this problem in the secondary schools.

Efforts with Pre-College Students

The MAA publication, The Math in High School You'll Need for College, has been widely circulated to guidance counselors and students in the schools. More than 320,000 copies have been requested from the national headquarters, a record for MAA publications. A few colleges and universities have prepared careful statements to secondary schools in their states describing the mathematics requirements in programs at their institutions and what constitutes appropriate secondary preparation for the required mathematics. One college has developed a review workbook in algebra and trigonometry which students can obtain at the time they are admitted and work on over the summer months before writing the mathematics placement test in the autumn.

One large midwestern state university is using equivalent forms of its placement exam to test high school students in their junior year and to make recommendations on senior math courses that, if successfully completed, will correct deficiencies before students come to the university. Early results from this project are promising.

The project Women and Mathematics (WAM) and Blacks and Mathematics (BAM), which place visiting lecturers in the schools to talk about the importance of early preparation in mathematics, have data showing significant increases in enrollment in college-preparatory courses in the schools that have been involved in their programs. Indeed the full program of the MAA's Committee on Secondary School Lectures is an important part of the profession's effort to influence the preparation of students in the schools.

The Committee has been impressed by mathematics courses developed in some departments for both elementary and secondary inservice teachers. There is evidence that teachers respond to opportunities to study more mathematics and that these offerings need to be expanded. The discontinuation of the NSF Institutes of the 60's has left secondary teachers with too little opportunity for the continued study of mathematics. A few departments are using undergraduate education majors as tutors in their remedial courses and regard this activity as an important part of their training. Usually the undergraduates are simultaneously enrolled in teaching seminars.

Characteristics of Effective Remedial Programs

Among the remedial programs which the Committee studied, there is considerable diversity in goals, course content, instructional format, and cost. However, the Committee has identified certain characteristics that contribute to successful programs and hopes that summarizing these will be helpful to departments that are attempting to improve these efforts. The most successful programs share these characteristics:

-- Involvement of regular faculty in the direction of the program and in the instruction

-- Institutional and departmental commitment to the program, shown both through reasonable financial support and through appropriate rewards for faculty involved

-- Requirement of strong commitment of time and effort on the part of students, with attempts to communicate to students the scope and demands of the programs

-- Program of study tightly structured, with the individual's pace determined by the instructor (not the student) in response to the student's ability and progress

-- Use of curricular materials often developed (or at least synthesized) by faculty in the program

-- Strong emphasis on problem solving, perhaps with calculators, and on involving students with demanding problems at an elementary level; a de-emphasis on vocabulary

-- Content which both responds to real world experiences and provides interesting elementary applications, and at the same time strives for conceptual understanding

-- Integration of mathematical ideas both horizontally (e.g., analytic geometry, algebra) and vertically (e.g., awareness that processes are repeated in different settings)

-- Effort to achieve some specific short range objectives (e.g., computational skills with rational expressions) and also to get started on some longer range goals (e.g., understanding the notion of function)

-- Awareness of the needs of different groups of students in the remedial audience (e.g., 18 year olds vs. returning adults, untaught students vs. unsuccessful students, anxious students vs. lazy students) and instructional strategies that provide for these differences

-- Emphasis on an informal instructional atmosphere that provides considerable moral support to the students and encourages peer interaction in the learning of mathematics

-- Enthusiasm among teachers and strong respect for the potential of the students

-- Ratio of students to teachers not greater than 30:1 in courses that include topics in arithmetic and elementary algebra

-- Provision for interaction between teacher and individual students

-- Careful attention to the training and supervision of inexperienced teachers

Answers to Some Particular Questions

The following questions have frequently been raised with the Committee.

Q 1. Are any schools successfully using computers for remedial instruction?

A 1. There are schools that are making good use of computers for course management and record keeping. However, we do not have an example where computers are being successfully used on a large scale to teach those courses designed to prepare students for college-level mathematics. At least one pilot project is underway, but materials and performance data are not yet available.

Q 2. Are there examples where large lecture is effective at the remedial level?

A 2. We know of several schools that have tried large lectures in their courses below the calculus, usually in combination with small tutorial sessions. Many have abandoned this format. We have not had any faculty defend its effectiveness at this level.

Q 3. What ways have departments found to reduce the cost of remedial instruction?

A 3. Wide differences in student backgrounds plus the common apprehension of students toward the learning of mathematics make remedial instruction costly. Done well, it requires the involvement of some experienced teachers and some small group instruction. Universities are commonly using teaching associates and undergraduate aids. Some colleges are using undergraduates as tutors, in particular, undergraduates who have successfully completed the remedial program. When these inexperienced teachers are carefully trained and are supervised by, and sometimes teamed with, faculty, there is evidence the results can be good. Some schools charge student extra fees to provide costly remedial courses.

Q 4. What are colleges doing about credit for these pre-college level courses?

A 4. There are colleges where remedial courses carry regular credit and there are colleges where remedial courses are non-credit. More common, however, is an arrangement where remedial courses provide credit but that credit is not applicable to degree requirements or at least to distribution requirements. This arrangement means that grades in remedial courses are part of the student's grade point average and that a remedial course is part of the student's total course load; in some state schools, it also qualifies remedial courses for state subsidy.

The Board of Governors of the MAA issued a statement on college credit for high school level mathematics at its meeting on August 20, 1979, which says in part:

"...Mathematics courses offered in college should be examined to determine the extent of their overlap with high school mathematics, and where that overlap is substantial, the course should not provide credit toward college graduation; but the students should be graded on their work, and the results should be included in computing grade point averages."

Q 5. What are the best curricular materials for the remedial level?

A 5. A few years ago most materials on the market for college remedial courses looked like accelerated school texts. Faculty at several schools, judging the available materials unsatisfactory, have been developing their own materials. Some of these are beginning to appear commercially and could be more appropriate for the instruction of college students. Diagnostic instruments to identify specific weaknesses of individual students are also under development. It may, however, be characteristic of good instruction at this level that instructors continue to write at least part of their own materials.

In recent years, attention in the departments have been focused largely on beginning level courses for under-prepared 18-year olds. There appears to be increasing awareness of the needs of adult, returning students who also need elementary mathematics but whose experiences are very different from those of traditional freshmen. We expect some materials for this audience to become available.

Q 6. How can we find out what other schools are doing with remedial instruction?

A 6. Many frustrated people have asked this question. There has not been an effective mechanism for the broad exchange of information among people working with instruction at this level. People starting new programs frequently feel they are re-inventing the wheel. People who feel they have something to say are unsure where to say it. The Committee feels that high priority for the Association needs to be providing for the exchange of information among people working with remedial instruction. Some section meetings this year did include remedial instruction as part of their programs.

The Committee is preparing a packet of information which will contain summaries of the best efforts at representative institutions (both instructional programs and programs in the schools) together with the names of people who can provide detailed information about these programs. This packet will be available from the MAA Headquarters for Section leaders and individuals responsible for starting new programs or improving old ones. It should be mentioned that a reasonably well-developed network does exist for persons interested in working with the math-anxious. Those who want more information may refer directly to Appendix II and the Resources section of Sheila Tobias' book, Overcoming Math Anxiety. Other networks have also been called to the attention of the Committee and we will attempt to index them in the information packet.

Recommendations to MAA

The Committee on Improving Remediation Efforts feels that the Association must recognize remedial instruction as an important responsibility of the mathematics community and must provide leadership to departments in seeing that this instruction is of the highest quality.

More specifically, the Committee recommends:

1. That the Association circulate this report to its members, in particular to department chairpersons, urging departments to provide appropriate instruction for students who are not prepared for college-level mathematics and adequate rewards to faculty for leadership in this instruction.

2. That the Association urge colleges and universities to develop ways of communicating to the schools the mathematics requirements of their institutions, that they make clear to students in the schools what mathematics preparation is necessary for students to enter adequately prepared, that where appropriate they provide courses in mathematics for in-service teachers.

3. That the Association provide in the program of national meetings for the exchange of ideas among teachers at the remedial level through swap sessions, panels, and workshops, and that Section leaders be encouraged to include various aspects of remedial instruction in their meetings.

4. That publications of the Association be used regularly to communicate information about the content and pedagogy of remedial programs in several institutions and about the results of some programs with secondary students

5. That remedial curricula become a specific responsibility of the Committee on the Undergraduate Program in Mathematics and that CUPM, working closely with the Committee on Two-Year Colleges, give continued attention to this level of instruction.

Many people have consulted with the Committee and have provided course materials and performance data. The Committee is greatly indebted to them and wishes to thank particularly the following colleagues who attended the San Antonio meetings for the purpose of talking with the Committee.

James H. Johnson, Norfolk State University
Frederick W. Keene, California State College
Jerrold Kleinsten, SUNY at Stony Brook
J. Michael Shaughnessy, Oregon State University
Charles N. Walter, Brigham Young University
Barbara Yanosko, Humbolt State College

Report prepared by the Committee on Improving Remediation Efforts in the Colleges

Robert Bumcrot
Philip M. Cheifetz
Ronald M. Davis
Donald M. Hill
Eleanor G. Jones
Joan P. Leitzel, Chairperson
Jean J. Pedersen

CALENDAR OF FUTURE MEETINGS

Sixty-first Summer Meeting, Pittsburgh, Pennsylvania, August 17–19, 1981.
Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Duquesne University, Pittsburgh, Pennsylvania, May 15–16, 1981.

EASTERN PENNSYLVANIA AND DELAWARE, Pennsylvania State University, Ogontz Campus, April 4, 1981.

FLORIDA, Bethune Cookman College, Daytona Beach, March 13–14, 1981.

ILLINOIS, Illinois State University, Normal, May 1–2, 1981.

INDIANA, Indiana University–Purdue University, Indianapolis, April 11, 1981.

INTERMOUNTAIN, Brigham Young University, Provo, Utah, April 10–11, 1981.

IOWA, Coe College, Cedar Rapids, April 24–25, 1981.

KANSAS, Benedictine College, Atchison, April 10–11, 1981.

KENTUCKY, Jefferson Community College, Louisville, April 3–4, 1981.

LOUISIANA–MISSISSIPPI, Friday–Saturday before February 20. Deadline for papers three months before meeting.

MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, William and Mary College, Williamsburg, Virginia, April 11, 1981.

METROPOLITAN NEW YORK, Lehman College, CUNY, May 2, 1981.

MICHIGAN, Oakland University, Rochester, May 1–2, 1981.

MISSOURI, Northwest Missouri State University, Maryville, April 10–11, 1981.

NEBRASKA, University of South Dakota, Vermillion, South Dakota, April 10–11, 1981.

NEW JERSEY, Seton Hall University, South Orange, March 1981.

NORTH CENTRAL, Mankato State University, Mankato, Minnesota, May 1–2, 1981.

NORTHEASTERN, Saturday before Thanksgiving and third week in June.

NORTHERN CALIFORNIA, University of Santa Clara, March 14, 1981.

OHIO, Miami University, Oxford, April 10–11, 1981.

OKLAHOMA–ARKANSAS, Oklahoma Christian College, Oklahoma City, March 27–28, 1981.

PACIFIC NORTHWEST, second Saturday in June. Deadline for papers six weeks before meeting.

ROCKY MOUNTAIN, Colorado College, Colorado Springs, May 1–2, 1981.

SEAWAY, Syracuse University, Syracuse, New York, April 10–11, 1981.

SOUTHEASTERN, University of Alabama, Birmingham, April 10–11, 1981.

SOUTHERN CALIFORNIA, California State University, Long Beach, March 7, 1981.

SOUTHWESTERN, New Mexico State University, Las Cruces, April 3–4, 1981.

TEXAS, San Antonio College, San Antonio, April 10–11, 1981.

WISCONSIN, University of Wisconsin, La Crosse, March 28–29, 1981.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.

AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES

AMERICAN MATHEMATICAL SOCIETY, Pittsburgh, Pennsylvania, August 18–21, 1981.

AMERICAN SOCIETY FOR ENGINEERING EDUCATION

ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.

ASSOCIATION FOR SYMBOLIC LOGIC

ASSOCIATION FOR WOMEN IN MATHEMATICS

CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/ SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHEMATIQUES

FIBONACCI ASSOCIATION

INSTITUTE OF MATHEMATICAL STATISTICS

MU ALPHA THETA

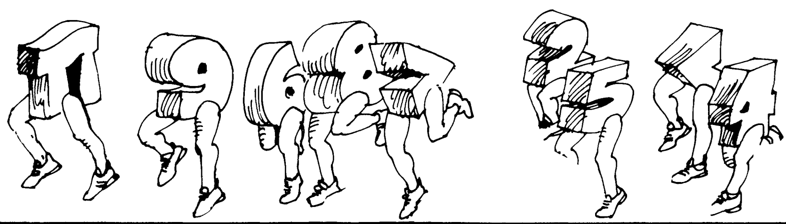
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, St. Louis, Missouri, April 22–25, 1981.

OPERATIONS RESEARCH SOCIETY OF AMERICA, Four Seasons Sheraton, Toronto, Canada, May 4–6, 1981.

PI MU EPSILON

SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.



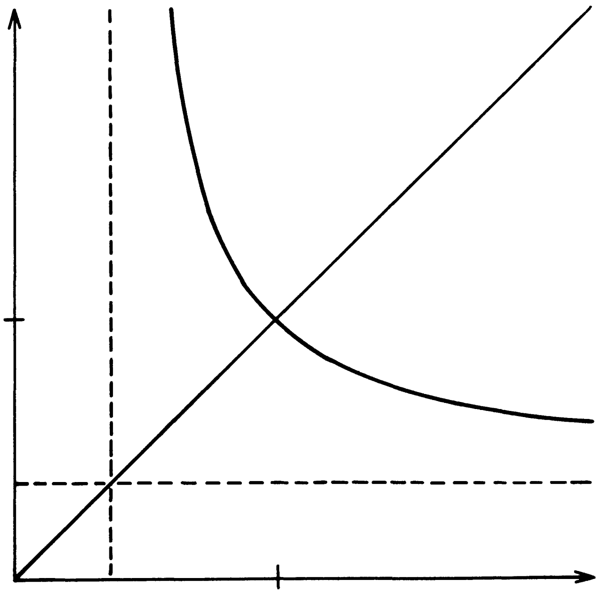
CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

APRIL

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 4



Iterated Exponentials

Old Theorems, New (and Old) Proofs:

Zeros of Polynomials

Quadratic Reciprocity

No Retraction

Shifted-Prime Factorizations

Fanout-Free Functions

$X^{X^{X^{X^{\cdot^{\cdot^{\cdot}}}}}}$

Detailed contents on cover 4

1981

THE AMERICAN
MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 4**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. (This is a change from the style previously used in this MONTHLY.) All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

EDITORIAL CORRESPONDENCE AND MAIN ARTICLES: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201; NOTES, etc.: to the corresponding Associate Editor; REPRINT PERMISSION: to R. P. BOAS (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

EXPONENTIALS REITERATED

R. ARTHUR KNOEBEL

Department of Mathematics, New Mexico State University, Las Cruces, NM 88003

Three questions motivate this article. When is x^y less than y^x? For what kind of numbers does $x^y = y^x$? And is there a formula for y as a function of x? The second question has attracted attention for some 250 years, since the time of Daniel Bernoulli's [28] interest in the integral solutions up to the present with the work of Sato [72] characterizing the algebraic numbers on the locus.

As regards the relationship between x^y and y^x, without apparent cause sometimes one, sometimes the other, is greater. For example,

$$1^2 < 2^1, \quad 2^2 = 2^2, \quad 3^2 > 2^3, \quad 4^2 = 2^4, \quad 5^2 < 2^5.$$

And recently Varner [76] proved a number of cases, for example, $x^e < e^x$ $(0 < x \neq e)$, among others. Is there a general pattern? Yes, but curiously many of the writers investigating this relation give no references whatsoever to previous work. It seems appropriate, therefore, in the hope of preventing further rediscovery, to review some of the most noteworthy facts about the commutativity of x^y, or the lack of it, and to cite the literature for the others that we don't touch on.

By graphing the equation

$$x^y = y^x, \tag{0.1}$$

we will split up the first quadrant into four regions according to the direction of inequality and thereby quickly answer any such questions about the relative magnitudes of x^y and y^x; this pictorial and historical approach even suggests methods of proof. Next we give the parametrization by Goldbach [29] of one of the curves separating these regions, on which Mahler & Breusch et al. [63] recently discovered the algebraic points.

We answer the third question by finding explicit expressions for y as a function of x. This leads us to the infinitely iterated exponential

$$x = h(z) = z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}.$$

The first thing to settle about this "function" is whether it ever converges, and if so, where? Surprisingly, it converges for all z in the interval $[e^{-e}, e^{1/e}]$ and diverges for any other positive z. With convergence established, we go on to find two expressions for y in terms of x for our original equation.

These facts about $h(z)$ for real z were first discovered and proved by Euler [78], Eisenstein [44], and Seidel [73], and then rediscovered by others many times over. However, when one turns to complex values, not much is known. So we will close this article with several open problems on convergence, analytic continuation, bifurcation, and general recursion.

Before beginning, let us say clearly what we mean by x^y. In order to avoid complications in defining exponentials on negative or complex numbers, let us declare at the outset that for the first three sections of this paper all our variables x, y, z are positive real numbers. The function x^y has a standard definition for integral arguments; by the use of root extraction it may be extended to the rationals; and with continuity, its values for real numbers may be inferred. Alternatively, in a more modern mode of definition,

$$x^y = e^{y \ln x}.$$

In either case, exponentiation is jointly continuous in both arguments.

The author received his Ph.D. under the direction of Alfred Foster at the University of California, Berkeley, in 1965. Since then he has been at New Mexico State University. His interests are in general algebraic systems and theoretical computer science. He became interested in the commutativity of powers when he was an undergraduate in Physics, and his interest was renewed when he accidentally came across Varner's article.—*Editors*

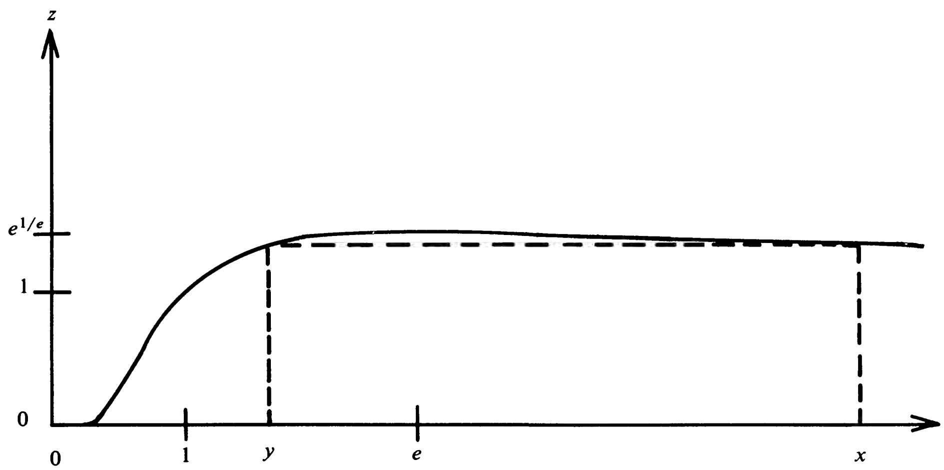


FIG. 1. $z = g(x) = x^{1/x}$

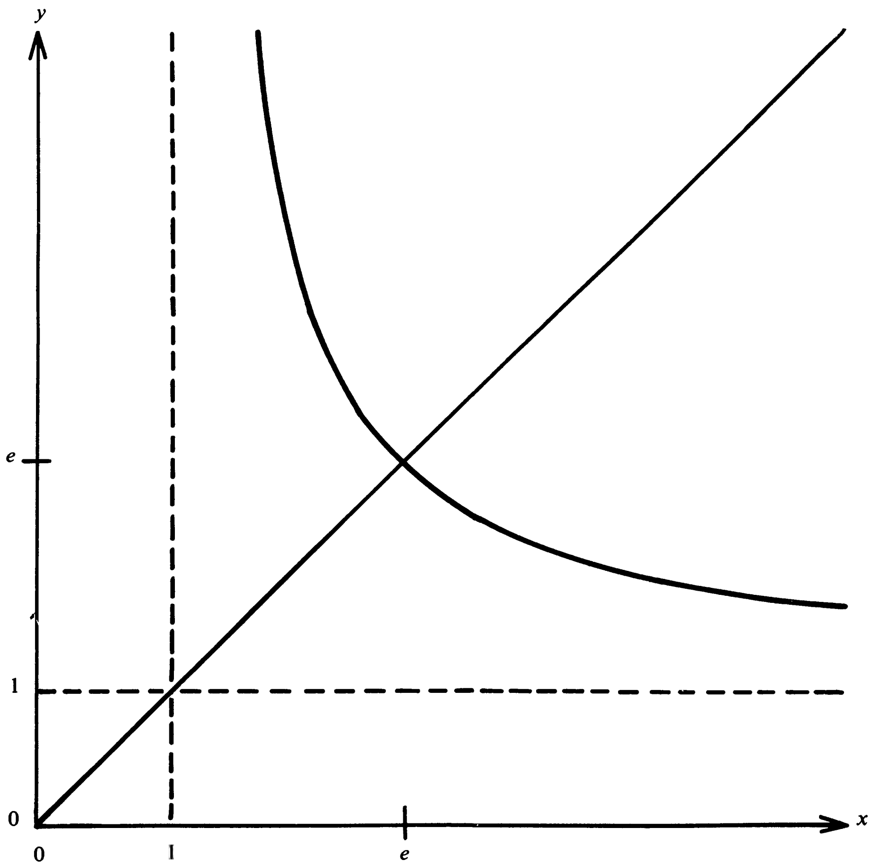


FIG. 2. $x^y = y^x$

For extending my expertise, many thanks are due to the referee, Ralph Boas, Carlton Evans, Michael Frese, Martha Gilmore, Hans Samelson, Ralph Wilkins, and many others who helped fill out the bibliography.

1. $x^y = y^x$. For convenience define

$$f(x,y) = x^y - y^x.$$

Now, for such values, $f(x,y)$ is a continuous function of both x and y. Therefore, to determine how its sign varies, we need only plot $f(x,y) = 0$ to find the boundary between regions of opposite sign. To this end, separate variables by taking xth and yth roots and obtain

$$x^{1/x} = y^{1/y}.$$

The auxiliary function $g(x) = x^{1/x}$ has the graph shown in Fig. 1. The salient points about this function, including the fact that $g(x)$ takes on its maximum $e^{1/e}$ at $x = e$, are given in Rotando & Plummer [77]. Now to plot $x^y = y^x$, simply find those abscissas which give equal values of $z = g(x)$. This yields Fig. 2 (Euler [48]). Notice that this locus divides the first quadrant into four regions alternating in the sign of $f(x,y)$.

The examples given earlier now emerge out of this picture. In particular, the fact that $x^e < e^x$ $(0 < x \neq e)$ corresponds in Fig. 2 to the fact that the line $y = e$ always manages to stay in the two negative regions. Any other line $y = y_0 \neq e$ would have to pass in part through a positive region.

The locus determined by $f(x,y) = 0$ has two branches intersecting at the point (e,e). The branch of equal points, $x = y$, we dismiss as trivial. Happily, the unequal branch can be parametrized

$$\begin{cases} x = s^{1/(s-1)}, \\ y = s^{s/(s-1)}, \end{cases}$$

by positive real numbers s (Goldbach [29]). Thus for each unequal pair (x,y) there is a unique $s = y/x$, the slope of the line from the origin to (x,y), which specifies that point. To check that these expressions really satisfy $x^y = y^x$ is a good exercise in the rules of exponentiation. Another exercise is to demonstrate Carlini's [89] relation, $xs = x^s$, which gives us simultaneously the solution of when the product and power functions are equal.

With this parametrization we can manufacture lots of simple pairs (x,y) such that $f(x,y) = 0$.

s	$1/2$	$2/3$	$4/3$	$3/2$	$5/3$	2	3	4
x	4	$27/8$	$64/27$	$9/4$	$(5/3)^{3/2}$	2	$3^{1/2}$	$4^{1/3}$
y	2	$9/4$	$256/81$	$27/8$	$(5/3)^{5/2}$	4	$3^{3/2}$	$4^{4/3}$

Note that if s gives the pair (x,y), then its reciprocal $1/s$ gives the converse pair (y,x). For $s = 1$, there is an apparent singularity, but it is only apparent. For if we take the limit $s \to 1$ and substitute $s = t + 1$, then we get the usual textbook expression defining e:

$$\lim_{t\to 0}(1+t)^{1/t} = x = e = y = \lim_{t\to 0}(1+t)^{1+1/t}$$

This parametrization easily characterizes certain important types of points. For example, as Euler [48] was the first to note, the points with rational coordinates are given by

$$s = \frac{m}{m+1} \quad \text{and} \quad s = \frac{n+1}{n}$$

for m and n positive integers.

The main point of the papers by Mahler & Breusch et al. [63] and Sato [72] is that the algebraic points (x,y) on the unequal branch are just those for which the parameter s is a rational m/n. They prove this by means of the Gel′fond-Schneider theorem (Baker [75]). Recall that a number is *algebraic* if it is the root of some polynomial with rational coefficients and a point is *algebraic* if both coordinates are algebraic numbers. In the case at hand, these algebraic

numbers are especially simple: they are just roots of rational numbers; that is, the polynomial equations they satisfy are

$$x^{m-n} = \left(\frac{m}{n}\right)^n,$$

$$y^{m-n} = \left(\frac{m}{n}\right)^m.$$

In the case when s is the fraction $m/n > 1$, and when $m/n < 1$, the inverses of both sides of these equations must be taken.

A subclass of the algebraic points is the class of points whose coordinates are algebraic integers: the polynomial equations must have integral coefficients with the leading coefficients being 1. In this class the parameter s simplifies to

$$m \quad \text{or} \quad \frac{1}{n}.$$

We summarize these three paragraphs in a table.

Type of point	s
Rational	$\frac{m}{m+1}, \quad \frac{n+1}{n}$
Algebraic	$\frac{m}{n} \quad (m \neq 1 \text{ or } n \neq 1)$
Algebraically Integral	$m, \quad \frac{1}{n}(m \neq 1 \text{ and } n \neq 1)$

Notice that there are only two points which are both rational and algebraically integral, namely, (2, 4) and (4, 2) for $s = 2$ and $1/2$; of necessity these are the only points on the curve with integral coordinates.

A closely related equation is

$$\bar{x}^{\bar{x}} = \bar{y}^{\bar{y}}.$$

One sees that this is equivalent to (0.1) by substituting inverses:

$$\bar{x} = \frac{1}{x},$$

$$\bar{y} = \frac{1}{y};$$

Fig. 5 is its graph. On logarithmic graph paper, the curve of Fig. 5 would be just the reflection through the point (1, 1) of the curve of Fig. 2. The previous analysis can be routinely carried over to this new curve with no surprises, except for this one thing: there are no algebraically integral points.

2. An Application in Biochemistry. Lest the reader think that what we have done so far is all pure, recreational mathematics with no possibility of application, we dash this illusion with an example from the study of enzyme reactions in biochemistry. Consider the situation of two substances, called substrates in enzyme kinematics, that can react individually with a common enzyme, which is assumed to be much less in quantity than either of the substrates, and whose activity, therefore, must be divided between the substrates. To be specific, suppose that we are hydrolysing d-ethyl mandelate and l-ethyl mandelate with pig's liver lipase. The respective products are d-mandelic acid and l-mandelic acid. The letters "d-" and "l-" in these names stand for dextrorotatory and levorotatory, which mean that polarized light is rotated to the right or left when passed through the corresponding products (and to a lesser extent in the corresponding substrates). If only one substrate were present, the extent of hydrolysis could be measured by the degree of polarization of the solution. However, when both are initially present in equal

quantities, only the difference in the magnitudes of the products can be so measured. Now there may be a differential action of the enzyme on the substrates with the result that one reaction races ahead of the other. This occurs at a maximum just when the difference of the products is maximal, and this in turn just when the optical rotation is a maximum. (Of course, if the reactions proceed at different rates, the substrates will also differ in concentration, but it is easy to see that this difference in substrate concentration is proportional to the difference in product concentration.)

Without going into the mathematical details (Haldane [30, pp. 85–88, 102–105] has an extensive exposition), the Michaelis-Menten theory tells us that this maximum occurs when

$$\bar{x} = s^{1/1-s}, \quad \bar{y} = s^{s/1-s};$$

where $\bar{x},\bar{y}$ are the current concentrations of the substrates, which are used up as the reaction proceeds, and s is a constant determined solely by the substrates and the enzyme. The reader will surely now recognize these expressions as the reciprocals $\bar{x} = 1/x, \bar{y} = 1/y$ of the parametrization given earlier of $x^y = y^x$. Thus the concentrations themselves satisfy

$$\bar{x}^{\bar{x}} = \bar{y}^{\bar{y}}.$$

For the example at hand, $s = 2$.

3. $z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}$. Can one do better than this parametrization of the last section and express y in an explicit closed form as a function of x? Well, there are explicit, if not finite, formulas for y. After all, if we can find the inverse of $z = g(x) = x^{1/x}$, then we can solve for y in the equation $y^{1/y} = x^{1/x}$. Consider the function (Euler [78])

$$y = h(z) = z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}$$

where the dots mean the limit of the sequence

$$z, z^z, z^{(z^z)}, \ldots$$

with the association of powers to the upper right. The function h is a partial but not complete inverse of g; even this much, though, will suffice to get a solution on one-half of the unequal branch. It surprises most people to learn that $h(z)$ converges for some $z > 1$. However, it does for all z running from e^{-e} up to $e^{1/e} \doteq 1.44$; see Fig. 3. (It is instructive for students to iterate powers on a pocket calculator and see the sequence slowly converge for $z = 1.4$ but blow up for $z = 1.5$ after an initial hesitation. In fact, I stumbled on this unexpected behavior of h while fooling around on a computer terminal using the interactive language APL, which invites just such experimenting because exponentiation is one primitive binary operation among lots of others and finite iteration is itself a primitive. See also Laidler & Landau [77], Wellen [78], and Wilson [77].)

Since the interval of convergence of h is somewhat unexpected (but was known by Euler [78]), we state this as a theorem. Two dichotomies split the statement and proof of the theorem into four parts according to which side of 1 z is on, and according to whether $h(z)$ converges or diverges.

For the convenience of the reader we give the numerical values of the various constants that will appear:

$$\begin{aligned} e^{1/e} &\doteq 1.444667861, & e^{-1/e} &\doteq .6922006276, \\ e &\doteq 2.718281828, & e^{-1} &\doteq .3678794412, \\ e^e &\doteq 15.15426224, & e^{-e} &\doteq .06598803585. \end{aligned}$$

It is also useful to have Maurer's [01] notation for "hyperpowers":

$${}^1z = z, \quad {}^2z = z^z, \quad {}^3z = z^{(z^z)}, \quad \ldots;$$

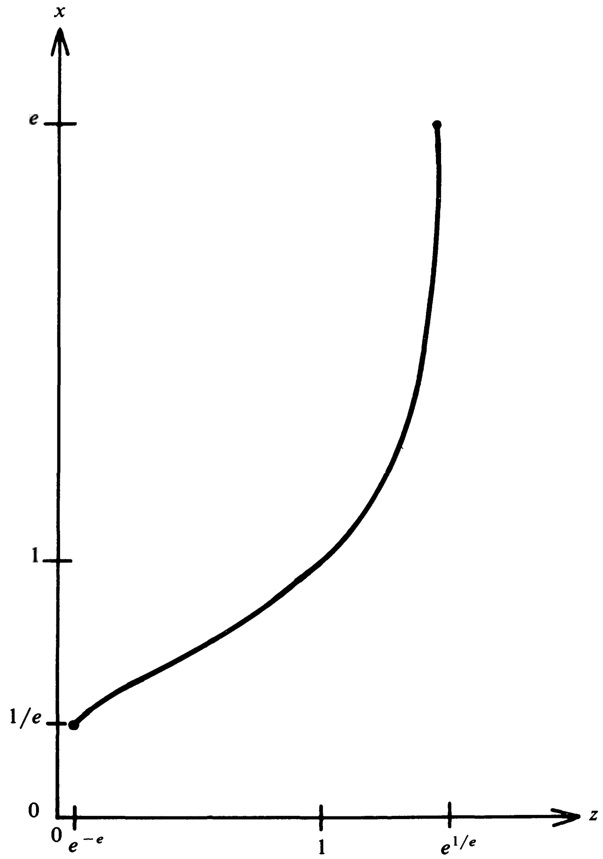


FIG. 3. $x = h(z) = z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}$

that is,

$$ {}^{n}z = \underbrace{z^{z^{\cdot^{\cdot^{\cdot^{z}}}}}}_{n \text{ times}} $$

THEOREM. *The function* $x = h(z) = z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}$ *converges when* $e^{-e} \leqslant z \leqslant e^{1/e}$ *and diverges for all other positive* z *outside this interval. On this interval* h *is the partial inverse of* g, *that is,*

$$ g(h(z)) = z \qquad (e^{-e} \leqslant z \leqslant e^{1/e}), $$
$$ h(g(x)) = x \qquad (e^{-1} \leqslant x \leqslant e). $$

In particular, four nontrivial modes of convergence and divergence occur.

Case 1: $z > 1$: *The sequence of hyperpowers increases monotonically*:

$$ z < {}^{2}z < {}^{3}z < \cdots. \tag{2.1} $$

Subcase 1c: $1 < z \leqslant e^{1/e}$. *The sequence* (2.1) *is bounded by* e, *and so* $h(z)$ *converges.*
Subcase 1d: $e^{1/e} < z$. *The sequence* (2.1) *increases without bound, and so* $h(z)$ *diverges.*

Case 2: $z < 1$: *The sequence of hyperpowers oscillates*:

$$ z < {}^{2}z > {}^{3}z < {}^{4}z > \cdots, $$

and the two subsequences

$$z < {}^3z < {}^5z < \cdots$$

and

$${}^2z > {}^4z > {}^6z > \cdots$$

each converge.

Subcase 2c: $e^{-e} \leqslant z < 1$. *The preceding two subsequences of odd and even hyperpowers converge to the same value, and so $h(z)$ converges.*

Subcase 2d: $z < e^{-e}$. *The preceding two subsequences each converge separately to different values, and so $h(z)$ diverges.*

Proof. Before starting, we remind the reader of three pertinent inequalities for positive a, b, and c that will occur repeatedly throughout the proof:

$$\text{if } a < b, \text{ then } a^c < b^c;$$

$$\text{if } a < b \text{ and } c < 1, \text{ then } c^a > c^b;$$

$$\text{if } a < b \text{ and } c > 1, \text{ then } c^a < c^b.$$

When the sequence of hyperpowers converges, that is, when $h(z) = x$ holds, then by reason of the continuity of exponentiation $z^x = x$ or, what is the same,

$$z = x^{1/x} = g(x).$$

From this follows the equations stating that h and g are inverses over the interval of convergence.

Case 1: $z > 1$: With the inequalities above, we easily obtain

$$z < {}^2z < {}^3z < \cdots.$$

Subcase 1c: $1 < z \leqslant e^{1/e}$. We prove convergence by showing that the sequence of hyperpowers is bounded above by e. By induction on the index of the hyperpowers, it suffices to verify that

$$\text{if } 1 < z \leqslant e^{1/e} \text{ and } w \leqslant e, \text{ then } z^w \leqslant e.$$

And this last follows from the inequalities just mentioned:

$$z^w \leqslant (e^{1/e})^e = e.$$

Thus the sequence converges since it is bounded and monotonically increasing.

Subcase 1d: $e^{1/e} < z$. As already seen, if the sequence of hyperpowers were to converge to x, then $z = x^{1/x} = g(x)$. But we have already shown that g has a maximum at $x = e$; hence $z = x^{1/x} \leqslant e^{1/e}$, and therefore h would be double valued, which is impossible (look at the graph of g, Fig. 1).

Case 2: $z < 1$. This case requires a subtler proof than the first since the sequence of hyperpowers alternates between two monotonically converging subsequences. This oscillation follows directly from the preliminary inequalities:

$$z < {}^2z > {}^3z < {}^4z > \cdots;$$

and so do these chains:

$$z < {}^3z < {}^5z < \cdots,$$

and

$${}^2z > {}^4z > {}^6z \cdots.$$

Realizing that ${}^nz$ is bounded, that is, $0 < {}^nz < 1$ for $0 < z < 1$, and therefore that these last two

sequences converge, we introduce two new functions, h_0 and h_e, defined for all positive $z \leqslant e^{1/e}$:

$$h_0(z) = \lim_{\substack{n\to\infty \\ n \text{ odd}}} {}^n z, \qquad h_e(z) = \lim_{\substack{n\to\infty \\ n \text{ even}}} {}^n z.$$

In case 1c already considered, it's clear that

$$h_0(z) = h_e(z) = h(z) \qquad (1 \leqslant z \leqslant e^{1/e}).$$

We now have to consider the real possibility that $h_0(z) \neq h_e(z)$ for some $z < 1$.

To find out how and why this split may occur, it is convenient to introduce the more involved relationship

$$z^{z^x} = x.$$

It is easy to see that if the point (z,x) satisfies any one of the relations

$$\begin{aligned} x &= h(z), \\ x &= h_0(z), \\ x &= h_e(z), \\ z &= g(x) \ \text{(or equivalently } z^x = x), \end{aligned}$$

then (z,x) also must satisfy

$$z^{z^x} = x.$$

As we proceed through the proof of Case 2, we will also gradually accumulate all the evidence necessary to establish the converse: namely, that the locus of the last equation is the union of the loci of the preceding four. To both these ends, write $j(z,x) = z^{z^x} - x$ and calculate the differential

$$\begin{aligned} dj(z,x) &= \frac{\partial j}{\partial z} dz + \frac{\partial j}{\partial x} dx \\ &= z^{z^x} z^{x-1}(x \ln z + 1)\, dz + (z^{z^x} z^x \ln^2 z - 1)\, dx. \end{aligned}$$

The coefficients of dz and dx are both simultaneously zero in the first quadrant just when

$$z = e^{-e} \quad \text{and} \quad x = e^{-1}.$$

Through any other point, by the implicit function theorem, there is a unique trajectory or line satisfying $j(z,x) = 0$. With this observation we can now settle the first subcase.

Subcase 2c: $e^{-e} \leqslant z < 1$. Since $h_0(1) = h(1) = h_e(1)$, it must be that $h_0(z) = h_e(z)$ for z down to (but not necessarily including) the critical value e^{-e}. (Equality at this critical value will follow from the next case.) Hence $h(z)$ converges to the common value.

Subcase 2d: $z < e^{-e}$. Divergence in this case is arrived at in three steps. First, we show for n even that ${}^n z \geqslant e^{-1}$. Then, from this it follows for n odd that ${}^n z \leqslant e^{-1}$. And finally we return to n even and establish that ${}^n z > e$ strictly.

Eisenstein [44] showed that the function z^{z^β} (β a constant) has a minimum $e^{-1/\beta e}$ at $z = e^{-1/\beta}$ (this can be established by elementary calculus). In particular, if $x \geqslant e^{-1}$, then $z^{z^x} \geqslant e^{-1}$. By induction, it will follow that for n even, ${}^n z \geqslant e^{-1}$. Raising z to these powers reverses inequalities, and so for $z < e^{-e}$ and n odd, we have

$${}^n z = z^{({}^{n-1}z)} < (e^{-e})^{e^{-1}} = e^{-1}.$$

Hence for $z < e^{-e}$, $h_0(z) \leqslant e^{-1} \leqslant h_e(z)$. Using the differential of $j(z,x) = z^{z^x} - x$ again, we see that $dx/dz > 0$ on the open line

$$L = \{\langle z,x \rangle | 0 < z < e^{-e} \text{ and } x = e^{-1}\}.$$

It must be then that $h_e(z) \neq e^{-1}$ when $0 < z < e^{-e}$, for if it were otherwise then, for

some z' in a neighborhood of this z, it would be that $h_e(z') < e^{-1}$, which is a contradiction. Q.E.D.

To fill out what was already mentioned in the proof about the union of all the loci being $j(z,x) = 0$, we add a corollary to the proof (see also Fig. 4).

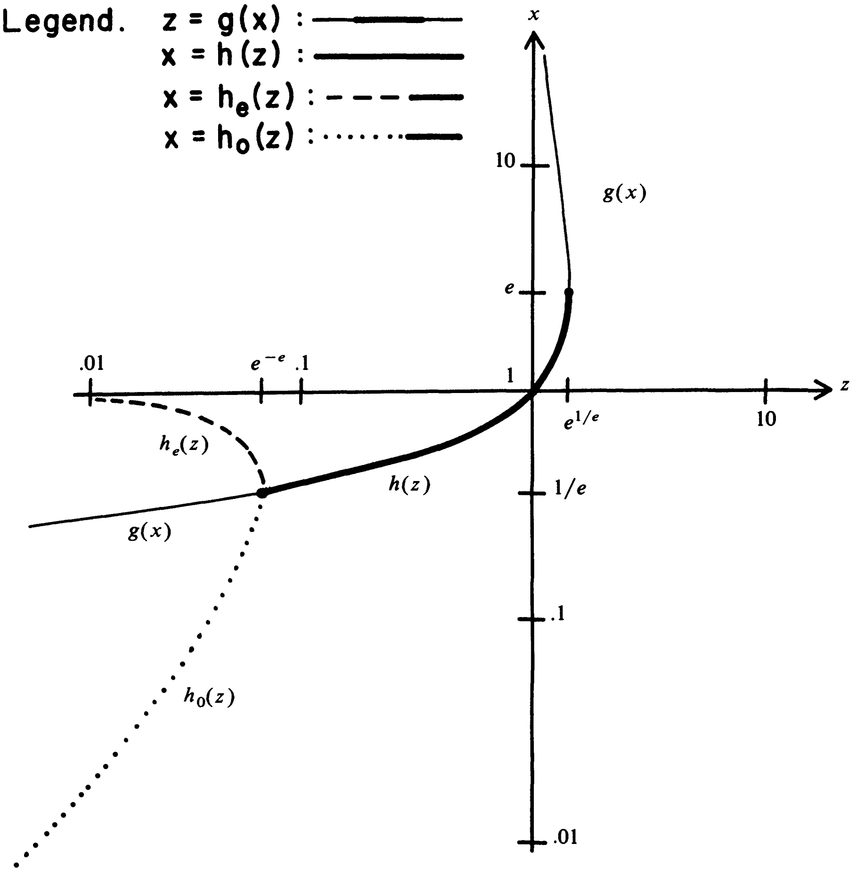


FIG. 4. $z^{z^x} = x$

COROLLARY. *The graph of the relation* $z^{z^x} = z$ *is exactly the union of the graphs of the functions* g, h, h_0, h_e *with the now established convention on the arguments*:

$$g(x) = z,$$
$$x = h(z),$$
$$x = h_0(z),$$
$$x = h_e(z).$$

Proof. As already noted, each of the individual loci satisfy $j(z,x) = 0$. For the other direction it suffices to show that there are no other points on $j(z,x) = 0$. And for this it suffices to establish that on each vertical line, i.e., $j(z_0, x)$ as a function of x alone, there are the right number of zeros. And this in turn will follow from the fact that the number of maxima and minima is exactly what it should be:

z_0	Number of extrema of $j(z_0, x)$	Number of zeros of $j(z_0, x)$
$0 < z_0 < e^{-e}$	2	3
$e^{-e} \leqslant z_0 \leqslant 1$	0	1
$1 < z_0 < e^{1/e}$	1	2

And finally the number of extrema of $j(z_0, x)$ is found as usual by calculating $\partial j/\partial x$, which we have already done, and setting it equal to zero.

As an aside, there is a bit of numerology in these bounds on z:

$$e^{-e} = (e^e)^{-1},$$
$$e^{1/e} = e^{(e^{-1})}.$$

The forms on the right are e raised to the additive and multiplicative inverses of e, and those on the left are the two different ways of associating a triple power (exponentiation is neither associative nor commutative).

But I digress. We must get back to expressing y in terms of x on the unequal branch of the relation $x^y = y^x$. This relation is equivalent to

$$g(y) = y^{1/y} = x^{1/x} = g(x).$$

Since we have just shown that h is the partial inverse of g, we obtain the desired result

$$y = h(g(y)) = h(x^{1/x}) = (x^{1/x})^{(x^{1/x})^{(x^{1/x})^{\cdot^{\cdot^{\cdot}}}}}$$
$$= x^{x^{-1+x^{-1+x^{-1+\cdot^{\cdot^{\cdot}}}}}} \qquad (x \geqslant e).$$

It might appear that we have simply applied a function to its inverse and so would get nothing new, but one should bear in mind that on the one half of the unequal branch for which $x > e$ we are out of the range of h. Thus we have started with x, passed to $z = g(x)$, and dropped back down to $y = h(z) = h(g(x))$ as already shown in Fig. 1.

In a similar manner one obtains an expression for half of the unequal branch of the locus in Fig. 5 of $\bar{x}^{\bar{x}} = \bar{y}^{\bar{y}}$. Thus

$$\bar{y} = \bar{x}^{\bar{x}^{1-\bar{x}^{1-\bar{x}^{1-\cdot^{\cdot^{\cdot}}}}}} \qquad (\bar{x} \leqslant 1/e).$$

There is another explicit form (although still not closed) for y as a function of x. From the theorem and the papers of Eisenstein [44] or Wittstein [45], we find

$$h(z) = 1 + \ln z + \frac{3^2(\ln z)^2}{3!} + \frac{4^3(\ln z)^3}{4!} + \cdots \qquad (e^{-1/e} < z < e^{1/e}). \tag{2.2}$$

Therefore, with $z = x^{1/x}$ and $y = h(z)$, it follows that

$$y = 1 + \frac{\ln x}{x} + \frac{3^2}{3!}\left(\frac{\ln x}{x}\right)^2 + \frac{4^3}{4!}\left(\frac{\ln x}{x}\right)^3 + \cdots \qquad (x > e) \tag{2.3}$$

on the unequal branch (see also Carmichael [08]).

There is a curious historical note that gives an indication of how rigor in mathematics was evolving over the time Euler [78], Eisenstein [44], and Seidel [73] wrote their papers. Euler used mainly numerical examples together with some algebraic manipulation to convince himself of the interval of convergence and the existence of bifurcation. Eisenstein, unaware of Euler's paper, was simultaneously both careful and careless in questions about convergence. He explicitly assumes that $0 < z < 1$ (in that paper α plays the role of our z), and tacitly implies that $h(z)$ converges throughout this interval without so much as a hint of a proof. On the other hand, with reasonable care, he establishes that the right side of (2.2) converges for $e^{-1/e} < z < e^{1/e}$. With his previous assumptions about $h(z)$ converging for $0 < z < 1$, he concludes that (2.2) holds just when $e^{-1/e} < z \leqslant 1$. However, with the help of our theorem, we can establish the full interval of convergence of (2.2). With regard to $h(z)$, we now know that Eisenstein's

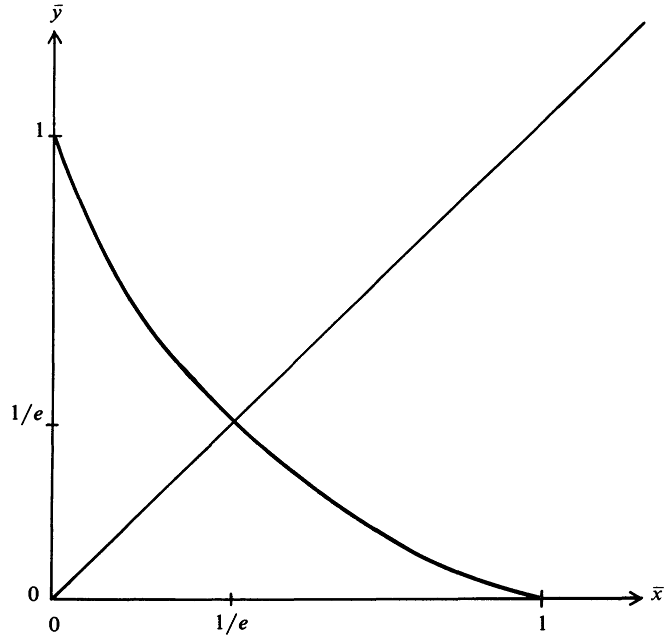


FIG. 5. $\bar{x}^{\bar{x}} = \bar{y}^{\bar{y}}$

implicit interval of convergence (0, 1) is wrong and must be stretched and moved over to the right to $[e^{-e}, e^{1/e}]$. To top this off, Seidel, also not knowing of Euler's paper, corrected Eisenstein's work and gave a complete proof for both convergence and bifurcation.

4. Open Problems. Despite the enormous literature on these topics when the variables are real numbers (see bibliography), once complex numbers are admitted, the literature becomes much, much smaller, and significant open problems appear. To limit the discussion, we look at only the iterated exponential. We describe four open problems about the region of convergence, the Riemann surface of analytic continuation, the nature of bifurcation, and higher levels of recursion.

Before proceeding we must cope with the fact that for complex values exponentiation is no longer single valued. By definition

$$z^w = e^{w \ln z};$$

its value depends on which sheet $\ln z = \operatorname{Log} r + i\theta$ is situated. Some authors get around this by working exclusively in terms of $t = \ln z$ as a new independent variable, but to maintain our present notation we agree to consider z as uniquely specified by its polar form $z = re^{i\theta}$. With this understanding, $\theta = 0$ and $\theta = 2\pi$ give distinct values of z.

Convergence. Where does $z^{z^{z^{\cdot^{\cdot^{\cdot}}}}}$ converge for complex z? On the one hand, Thron [57] established

$$|\ln z| \leqslant e^{-1}$$

as a region R_T of convergence. On the other hand, Carlsson [07] earlier showed that there is convergence only if $z = e^{\zeta e^{-\zeta}}$ for some ζ such that $|\zeta| \leqslant 1$; call this region R_C.

What about the points between R_T and R_C? Does the iterated exponential converge there? Shell [62] established several successively larger dumbbell-shaped regions of convergence which overlap that of Thron, leaving part of R_T uncovered along and near the real axis but extending well beyond R_T in other directions. However, the union of all these regions of established convergence fails to exhaust R_C itself. Thus our open problem is this: prove or disprove the

conjecture of Shell [59], based on considerable computer calculation, that the sequence $z, {}^2z, {}^3z, \dots$ converges for all $z \in R_C$.

Analytic continuation. An *analytic function* is one which can be expanded locally in a power series (see Churchill et al. [74]). For example, the series (2.2) shows that h is analytic in the interval $(e^{-1/e}, e^{1/e})$. It may be necessary to use more than one power series to represent an analytic function at various points. Specifically, we shall see shortly that h is analytic also in its whole domain of definition $(e^{-e}, e^{1/e})$ but the series (2.2) does not converge to the right of $e^{-1/e}$: there other power series are needed. The *analytic continuation* of a function along an arc going from its domain of definition and beyond it is a new analytic function defined on the whole arc and agreeing with the original function wherever it was defined. For our situation, we will show that h has an analytic continuation all the way down to (but not including) $z = 0$; it is in fact the inverse of g (see Fig. 4).

A fundamental result in the theory of complex variables is that analytic continuation is unique when it exists. But this is guaranteed only along the same arc: continuing the function $\sqrt{z}$ around the origin to the starting point will give the value $-\sqrt{z}$ of opposite sign. Thus analytic continuation may result in the complex domain being separated into two or more sheets. The analytic continuation of $\sqrt{z}$ has two sheets; the analytic continuation of $\ln z$ has an infinite number of sheets. This last fact will probably mean that the analytic continuation of h has many sheets (remember that $z^w = e^{w \ln z}$ is multivalued).

How much do we know about the analytic continuation $\hat{h}$ of h? We have already seen that if $\lim {}^nz$ converges to x then $z^x = x$ and so $z = x^{1/x} = g(x)$. (Here x is a complex variable and *not* the real part of z.) As an elementary function, $g(x)$ is analytic everywhere except at $x = 0$. It is also invertible everywhere except where $g'(x) = 0$, that is, at $x = e$. Hence its inverse is analytic except when $z = 0$ or $z = e^{1/e}$. Thus $h(z)$, as a segment of the inverse of $g(x)$, is itself analytic. There is an arc between any two x's other than 0 and e, so there is an analytic continuation between the corresponding z's. Hence all of the inverse of g is an analytic continuation of h.

In particular, we can travel from $x = 2$ to $x = 4$ along the real axis except at $x = e$, where we must take a slight detour around this point into the complex numbers because $g'(e) = 0$. This means that in the analytic continuation of h, we must be on different sheets:

$$\hat{h}(\sqrt{2}\,) = 2,$$

$$\hat{h}(\sqrt{2}\,) = 4.$$

This explains the anomaly of $2 = 4$ puzzled over by Etherington [38], Gottlieb [73], and Andrews & Lacher [77].

Analytic continuation explains another curiosity. From his famous equation $i = e^{i\pi/2} = (e^{\pi/2})^i$, Euler [78] also came across the solution $z = e^{\pi/2}$, $x = i$. Again this means that $\hat{h}(e^{\pi/2}) = i$ on some sheet in the analytic continuation. It is easy to see that on another sheet also $\hat{h}(e^{\pi/2}) = -i$.

If one is to solve this open problem of determining precisely the extent of the complete analytic continuation of h, one must check that the term $\ln z$ in the definition of h does actually introduce an infinite number of sheets, find all the additional sheets introduced by the lack of one-to-one-ness in its inverse g, and finally investigate the possibility that $x = h(z)$ might continue analytically beyond the relation $z^x = x$.

Bifurcation. Just as the branch g beyond the region of convergence of $\lim {}^nz$ in Fig. 4 is the germ for studying the analytic continuation of h, so the other two branches h_0 and h_e lead us to settle rather for convergence of subsequences in the complex plane. There are complex z with almost periodic behavior of the sequence

$$z, {}^2z, {}^3z, \dots,$$

but the periods are often different from 2 as in the real case, and the behavior is much more complicated. These three values of $z = re^{i\theta}$ have the following near periods:

r	θ	near period
4	3	3
3	4	4
3.5	3.5	5

The reader is encouraged to try his hand at finding other periods with, say, a programmable pocket calculator or a computer terminal.

The author is unable and unqualified to make head or tail of this chaotic behavior. Hence the third open problem is to explain the bifurcation of iterated exponentials on complex numbers. To get started see Straffin [78] or Marsden [78].

Recursion. Finally, there is a circle of problems centered around the iteration of hyperpowers

$$^{\cdot^{\cdot^{z^{z}}}}z.$$

That this makes any sense at all depends on the extension of ^{n}m from positive integers to at least all positive real numbers (see the note and references to Wright [47] and others in the bibliography). For infinite iteration, we can repeat the three open problems above about convergence, analytic continuation, and chaotic behavior.

For finite iteration, we can "hype" the hyperpowers again,

$$^{z^{\cdot^{\cdot^{z}}}\text{times}}z,$$

and then again, and away we go to any level of recursion. This has actually been done for the positive integers and the result is the three-argument Ackermann [28] function $\alpha(k,m,n)$, whose definition is given by

$$\begin{cases} \alpha(1,m,n) = m+n \\ \alpha(k+1,m,1) = m, \\ \alpha(k+1,m,n+1) = \alpha(k,m,\alpha(k+1,m,n)). \end{cases}$$

For the first four values of k, we get the familiar functions

$$\begin{aligned} \alpha(1,m,n) &= m+n \\ \alpha(2,m,n) &= m\cdot n, \\ \alpha(3,m,n) &= m^n, \\ \alpha(4,m,n) &= {}^{n}m. \end{aligned}$$

This indexing fits in well with the historical naming of hyperpowers as the "fourth natural algorithm" by Lémeray [98] and the "operation of the fourth kind" by Schubert [98].

The Ackermann function is the classic example in the theory of computation of a general recursive function which is not primitive recursive (see Hennie [77] and Calude, Marcus & Tevy [79]). The essential difference distinguishing these two kinds of algorithms is that a primitive recursive function, such as exponentiation, is defined by recursion on just one argument at a time, whereas a general recursive function may be defined by recursion on two or more arguments simultaneously, such as in the definition of Ackermann's function. This difference between double and single recursion is dramatically illustrated by the computer programs for α and any one of the binary arithmetical operations, say m^n: it takes incredibly more time and storage to compute, for example, $\alpha(3,5,5)$ as opposed to 5^5, even though the answer is the same.

General recursion goes well beyond the kind of definition which Wright [47] and others extended to the complex numbers. This leads to the final question: Can Ackermann's function be analytically continued to all numbers in the complex plane?

Bibliography.

Those who cannot remember the past are condemned to repeat it.—George Santayana

Here are some notes prefacing the references in order to help readers find their way through the literature. There are, first, additional references on the equation $x^y = y^x$, then some citations for the iterated exponential, and finally a miscellanea of notes about related papers. For the reader who wishes only to sample this extensive literature, the following accessible items are recommended: for the commutativity of exponentiation, Hausner [61], Hurwitz [67], Mahler & Breusch [63], and Moulton [16]; and for iterated exponentiation, Bender & Orszag [78], Barrow [36], Creutz & Sternheimer [?], Euler [78], Shell [62], Thron [57], and Wright [47].

Contrary to the initial observation that most articles in the area have few or no references, the relatively recent articles by Hausner [61] and Hurwitz [67] are exceptional in this respect and should be consulted for numerous citations on $x^y = y^x$. Likewise Dickson [19] and Archibald [21] should be consulted for older ones.

We have deliberately restricted ourselves to the case when both x and y are positive in this equation, since otherwise, to be modern, one should go over to complex numbers, which only Schwering [78] and Moulton [16] have completely done, so far as I know. The old-fashioned way of real root extraction, when it exists, is followed by Wittstein [45], Carmichael [08], Nesbitt [13], and Hurwitz [67]. Nesbitt [13] and Moulton [16] have nice graphs of $x^y = y^x$ for all real values of x and y, both positive and negative.

There is a related problem, which we have not mentioned, of when $x^y - y^x$ is not zero but equal to a prescribed integer. See Cassels [60], Krishnasastry & Perisastri [65], Schinzel [67], and Tijdeman [76] for special cases of this and also for additional references.

For the iterated exponential the most extensive surveys are to be found in Carlsson [07] and Shell [59]. Many authors have studied iterated exponentials with unequal exponents,

$$z_1^{z_2^{z_3^{\cdot^{\cdot^{\cdot}}}}},$$

for example, Barrow [36], Thron [57], and Creutz & Sternheimer [?]. This leads to the expansion of arbitrary numbers and functions by infinite exponential powers, which is discussed by Ditor [78] and Bender & Orszag [78].

Closely connected with the infinite iteration is the extension of the finite iterates to nonintegral values; in short, how should the hyperpower ${}^nx$ be defined for arbitrary n? For some selected references on this special problem as well as the more general one of iterating an arbitrary function and extending it to the reals, see Abel [?], Collins [38], Cayley [60], Schröder [71], Koenigs [84], Lemeray [95], Hadamard [44], Pincherle [06], Carlsson [07], and Wright [47], especially the last three for surveys and extensive bibliographies.

The mathematics of this article has been a fruitful source of problems: the equation $x^y = y^x$ appears in Bush [61], Mahler & Breusch [63], and Bryant et al. [65]; the iterated exponential in Lense [24], Bromwich [26], Francis & Littlewood [28], Chaundy [35], Knopp [51], Apostol [57], and Bryant et al. [65].

Just as the lack of commutativity in taking powers motivated this article, so the lack of associativity might also pique one's interest. However, the condition that

$$w^{(x^s)} = (w^x)^s,$$

when w, x, s are positive and $w \neq 1$, is equivalent to Carlini's [89] multiplicative relation $xs = x^s$, which in turn reduces to the trivial condition

$$x = s^{1/(s-1)},$$

already studied in §1. For the association of an arbitrary number of terms, see the contributions

of Wöpcke [51], P. Goodstein [58], R. L. Goodstein [58], Göbel & Nederpelt [71], Riordan [73], and Gardner [73].

To see some of the equations of this paper solved in terms of ordinals and cardinals, consult G. Cantor [97], Sierpiński [58], and Hickman [76]. (The equation $z^x = x$ plays a significant role in the theory of ordinals.)

Lastly, we mention here some historical details not covered before about who did what when. The comparison of x^y and y^x apparently goes back to Hengel [88]. The finite iteration of the exponential can be traced to Condorcet [78] and the infinite to Euler [78]. Wittstein [45] was the first to use $g(x) = x^{1/x}$ in connection with $x^y = y^x$, and Euler [78] first applied it to the study of $h(z)$. Of course, the logarithm of $g(x)$ goes back to Euler's formulation of the prime number theorem.

The MR numbers at the end of some of the citations refer to abstracts in the Mathematical Reviews: those of JFM to the Jahrbuch über die Fortschritte der Mathematik, wherever known.

References

Abel, Niels Henrik [?]. Oeuvres complètes, vol. 2, art. VI, pp. 36–39. Christiania (Oslo), Grøndahl & Søn, 1881. This article was published posthumously; the original date of writing is not known.

Ackermann, W. [28]. Zum Hilbertschen Aufbau der reellen Zahlen. Mathematische Annalen 99 (1928), pp. 118–133.

Allen, Arnold O. [69]. e^π or π^e? J. Recreational Math. 2 (1969), pp. 255–256.

Andrews, J. J.; Lacher, R. C. [77]. Stacked exponents. Aequationes Math. 16 (1977), pp. 137–147.

Apostol, Tom M. [57]. Mathematical Analysis. Addison-Wesley, 1st ed., 1957, p. 383, Exercise 12-7.

Archibald, Raymond Clare [21]. Problems—Note no. 9. This MONTHLY 28 (1921) pp. 141–143.

Baker, Alan [75]. Transcendental Number Theory. Cambridge University Press, 1975, pp. 10–11.

Barrow, D. F. [36]. Infinite exponentials. This MONTHLY, 43 (1936), pp. 150–160.

Barshinger, Morris H., Jr. [73]. A look at $x^{x^{\cdot^{\cdot^{\cdot}}}}$. J. Undergraduate Math. 5 (1973), pp. 37–41.

Bender, Carl M.; Orszag, Steven A. [78]. Advanced Mathematical Methods for Scientists and Engineers. McGraw-Hill, 1978. pp. 147–148, 196, 399–400, 414.

Bernoulli, Daniel [28]. Letter to Goldbach, June 29, 1728. Correspondance Mathematique et Physique de quelques célèbres géomètres du XVIII siècle (Paul Heinrich von Fuss, editor), vol. 2 (1843), p. 262.

Brodetsky, S. A First Course in Nomography. G. Bell and Sons, London, 1925, pp. 129–133. The author gives nomographic solutions to $a^x = x^b$ and related equations.

Bromwich, T. J. I'a [26]. An Introduction to the Theory of Infinite Series. Macmillan, 2nd ed., 1926. Example 11, pp. 23–24 and Exercise 4, pp. 160–161.

Bryant, Steven J.; Graham, George E.; Wiley, Kenneth G. [65]. Nonroutine Problems in Algebra, Geometry, and Trigonometry. McGraw-Hill, 1965. Problems 28 and 101, pp. 6, 15, 37–38.

Bush, L. E. [61]. The 1960 William Lowell Putnam Mathematical Competition. This MONTHLY 68 (1961), pp. 762–763.

Calude, C.; Marcus, S.; Tevy, I. [79]. The first example of a recursive function which is not primitive recursive. Historia Mathematica 6 (1979), pp. 380–384.

Cantor, Georg [97]. Beiträge zur Begründung der transfiniten Mengenlehre. Mathematische Annalen 49 (1897), pp. 207–246, especially pp. 242–246.

Cantor, M. [77]. About $x^y = y^x$. Verhandlungen der 32. Versammlung Deutscher Philologen und Schulmänner in Wiesbaden 1877. Teubner, Leipzig, 1878, pp. 179–180.

Carlini, Luigi [89]. Sull'uguaglianza $a^b = b^a$ con a e b interi e positivi. Periodico di Matematica 4 (1889), pp. 117–119.

Carlsson, Albert [07]. Om Itererade Funktioner. Ph.D. dissertation, Uppsala, 1907.

Carmichael, R. D. [06a]. Problem #211. If $x = v^{1/(v-1)}$, what is the $f(x)$ such that $v = f(x)$? This MONTHLY 13 (1906), pp. 18, 72.

Carmichael, R. D. [06b]. On a certain class of curves given by transcendental equations. This MONTHLY 13 (1906), pp. 221–226.

Carmichael, R. D. [08]. On certain transcendental functions defined by a symbolic equation. This MONTHLY 15 (1908), pp. 78–83.

Cassels, J. W. S. [60]. On the equation $a^x - b^y = 1$. II. Proc. Cambridge Philos. Soc. 56 (1960), pp. 97–103. Corrigendum, ibid. 57 (1961), p. 187. MR22, #5610.

Cayley, Arthur [60]. On some numerical expansions, Quarterly J. of Pure and Applied Math. 3 (1860), pp. 366–369. Also in Collected Math. Papers. Cambridge. 4 (1891), pp. 470–472.

Chaundy, T. [35]. The Differential Calculus. Oxford, 1935, pp. 408–409.

Churchill, Ruel V.; Brown, James W.; Verhey, Roger F. [74]. Complex Variables and Applications. McGraw-Hill, 3rd ed., 1974. Chapters 6 and 12.

Collins, Eduard [38]. Nouvelles recherches sur la théorie des puissances fonctionnelles. Mémoires de l'Académie des Sciences de St. Pétersbourg. 3 (1838), pp. 313–331. Usually listed now under Akademiia Nauk SSSR, Leningrad.

Condorcet, M.J.A.N.C. Marquis de [78]. Sur quelques séries infinies dont la somme peut être exprimée par des fonctions analytiques d'une forme particulière. Acta Academiae Scientiarum Petropolitanae 1 (1778), pp. 34–37.

Creutz, Michael; Sternheimer, R. M. [?]. On the convergence of iterated exponentiation. Fibonacci Quarterly (to appear).

Cunningham, Allan [16]. Factorisation of $N = (x^y \mp y^x)$. Messenger of Mathematics 45 (1916), pp. 185–192.

Dainelli, Ugo [89]. Sull'equazione $x^y = y^x$ con x ed y interi e positivi. Periodico di Matematica 4 (1889). pp. 115–117.

Dickson, Leonard E. [19]. A History of the Theory of Numbers. Carnegie Institution of Washington, 1919, vol. 2, p. 687.

Ditor, Seymour Z. [78]. Some binary representations of real numbers and odd solutions of the functional equation $\phi \circ f = g \circ \phi$. Aequationes Math. 18 (1978), pp. 265–266.

Eisenstein, G. [44]. Entwicklung von $\alpha^{\alpha^{\alpha^{\cdot^{\cdot^{\cdot}}}}}$. Crelle's Journal für die reine und Angewandte Mathematik 28 (1844), pp. 49–52.

Etherington, I. M. H. [38]. Mathematical Note 1323. 2 = 4: a new proof. Math Gazette 22 (1938), p. 403.

Euler, Leonhard [48]. Introductio in analysin infinitorum. 1748, reprinted by Culture et Civilization, Brussels, 1967, vol. 2, pp. 293–295.

Euler, Leonhard [78]. De formulis exponentialibus replicatis. Acta Academiae Scientiarum Petropolitanae 1 (1778), pp. 38–60. Also in Opera Omnia, Series Prima, vol. 15. pp. 268–297.

Flechsenhaar, A. [11]. Ueber die Gleichung $x^y = y^x$. Unterrichtsblätter für Math. 17 (1911), pp. 70–73.

Francis, E. C.; Littlewood, J. E. [28]. Examples in Infinite Series with Solutions. Deighton Bell, Cambridge, 1928, pp. 10, 38–39.

Franklin, Philip [17]. Relating to the real locus defined by the equation $x^y = y^x$. This MONTHLY 24 (1917), p. 137.

Gardner, Martin [73]. Mathematical games. Scientific American 228 (May 1973), p. 105, and (June 1973), p. 110.

Gardner, Martin [77]. Mathematical games—in which joining sets of points of lines leads into diverse (and diverting) paths. Scientific American 237 (1977), pp. 18ff, especially p. 28. The author explains Knuth's notation for the Ackermann function and its role in describing an extraordinarily large bound on a Ramsey number in graph theory.

Göbel, F.; Nederpelt, R. P. [71]. The number of numerical outcomes of iterated powers. This MONTHLY 78 (1971), pp. 1097–1103. MR 45, #6643.

Goldbach, Christian [29]. Reply to Daniel Bernoulli. Correspondance Mathématique et Physique de quelques célèbres géomètres du XVIII siècle (Paul Heinrich von Fuss, editor) vol. 2, (1843), pp. 280–281. Cited in Archibald [21].

Goodstein, Peter [57]. The iterated exponential of x. Math. Gazette 41 (1957), p. 219.

Goodstein, Peter [58]. Limits of iterated logarithmic functions. Math. Gazette 42 (1958), pp. 295–296.

Goodstein, R. L. [58]. The system of equations $b^x = x_1, b^{x_1} = x_2, \ldots, b^{x_n} = x$. Math. Gazette 42 (1958), pp. 296–299.

Gottlieb, Allan J. [73]. Given that $x^{x^{x^{\cdot^{\cdot^{\cdot}}}}} = 2$, find x. Technology Review, May 1973, Problem MAY2 p. 58. Solution by R. Robinson Rowe, ibid., Oct./Nov. 1973, pp. 71–72.

Gravé, D. A. [98]. Sur les expressions dites surpuissances. Nouvelles Annales de Mathématiques (3) 17 (1898), pp. 80–91. JFM 29, p. 343.

Hadamard, J. [44]. Two works on iteration and related questions. Bulletin of the American Math. Society 50 (1944), pp. 67–75.

Haldane, J. B. S. [30]. Enzymes. 1930. Reprinted by M.I.T. Press, Cambridge, Mass., 1965, pp. 85–88.

Hausner, Alvin [61]. Algebraic number fields and the Diophantine equation $m^n = n^m$. This MONTHLY 68 (1961), pp. 856–861. MR 24, #A1879.

Heis, Eduard [88]. Sammlung von Beispielen und Aufgaben aus der allgemeinen Arithmetik und Algebra. Verlag der M. DuMont-Schauberg'schen Buchhandlung, Cologne, 1888, pp. 371–372. Here is an assortment of numerical problems with exponentials.

Hengel, J. van [88]. Beweis des Satzes, das unter allen reellen positiven ganzen Zahlen nur das Zahlenpaar 4 und 2 für a und b der Gleichung $a^b = b^a$ genügt. Progr. Emmerich, 1888. Cited in Dickson [19].

Hennie, Fred [77]. Introduction to Computability. Addison-Wesley, 1977. Chapter 5.

Hessel [50a]. Ueber die Bedingung, unter welcher $a^x > x$ ist. Archiv der Mathematik und Physik 14 (1850), pp. 93–96.

Hessel [50b]. Ueber das merkwürdige Beispiel einer zum Theil punctirt gebildeten Curve, das der Gleichung entspricht: $y = \sqrt[x]{x}$. Archiv der Mathematik und Physik 14 (1850) pp. 169–187.

Herbst, C. [09]. Bestimmung der ganzen Zahlen, für welche die Beziehung gilt: $x^y = y^x$. Unterrichtsblätter für Mathematik und Naturawissenschaften 15 (1909), pp. 62–63.

Hickman, J. L. [76]. Analysis of an exponential equation with ordinal variables. Proceedings of the American Math. Society 61 (1976), pp. 105–111.

Hurwitz, Solomon [67]. On the rational solutions of $m^n = n^m$ with $m \neq n$. This MONTHLY 74 (1967), pp. 298–300. MR35, #127.

Isaacson, F. de st. Q. [52]. A note on multiple exponentials. Math. Gazette 36 (1952), pp. 127–128.

Klambauer, G. [79]. Problems and Propositions in Analysis. Dekker, 1979, pp. 186–193, 450–453.

Knopp, K. [51]. Theory and Application of Infinite Series. Blackie and Son, London, 1951. Exercise 23, p. 108.

Koenigs, G. [84]. Recherches sur les intégrales de certaines équations fonctionnelles. Annales Scientifiques de l'Ecole Normale Supérieure (3) 1 (1884), pp. 3–41.

Krishnasastry, M. S. R.; Perisastri, M. [65]. On some diophantine equations. Math. Student 33 (1965), pp. 73–76. MR33, #4006.

Laidler, P.; Landau, B. V. [77]. A power sequence exercise for a pocket calculator. Math Gazette 61 (1977), pp. 191–194.

Łazowski, T. [12]. On the convergence of certain numerical sequences. (Polish) Wektor 1 (1912), pp. 277–292. JFM 44, pp. 126–127.

Lecointe [52]. Solution de la question 235: Résoudre en nombres rationnels l'équation $x^y = y^x$. Nouvelles Annales de Mathématiques 11 (1852), pp. 187–189.

Lémeray, E. M. [95]. Sur les fonctions itératives et sur une nouvelle fonction. Association Française pour l'Avancement des Sciences, Congres Bordeaux 2 (1895), pp. 149–165.

Lémeray, E. M. [96]. Sur les racines de l'équation $x = a^x$. Nouvelles Annales de Mathématiques (3) 15 (1896), pp. 548–556 and 16 (1897), pp. 54–61.

Lémeray, E. M. [98]. Sur un mémoire d'Euler—Correspondance. Nouvelles Annales de Mathématiques (3) 17 (1898), p. 139. JFM 29, p. 343.

Lémeray, E. M. [98]. Le quatrième algorithme naturel. Proceedings of the Edinburgh Math. Society 16 (1898), pp. 13–35 + 8 figures. JFM 29, p. 343.

Lense, J. [24]. Problem 3053. This MONTHLY 31 (1924) pp. 500–501.

Lévy, Paul [27]. Sur l'itération de la fonction exponentielle. Acad. Sci. Paris. Comptes Rendus Heb. Seances. Académie de Sciences 184 (1927), pp. 500–502.

Littlewood, J. E. [48]. Large numbers. Math. Gazette 32 (1948), pp. 163–171.

Lowerre, George F. [79]. A logarithm problem and how it grew. The Math. Teacher 72 (1979), pp. 227–229. This is about $z^x = x$.

Luxenberg, M. [81]. Ueber die gleichung $x^y = y^x$. Archiv der Mathematik und Physik 66 (1881), pp. 332–334.

Macintyre, A. J. [66]. Convergence of $i^{i^{i}}$ Proc. Amer. Math. Soc. 17 (1966), p. 67. MR32, #5855.

Mahler, K.; Breusch, Robert; et al [63]. Problem 5101: Solutions of an old equation. This MONTHLY 70 (1963), p. 571, statement of problem. Ibid. 71 (1964), p. 564, solution.

Marsden, Jerrold E. [78]. Qualitative methods in bifurcation theory. Bulletin of the American Math. Society 84 (1978), pp. 1125–1148.

Maurer, Hans [01]. Über die Funktion $y = x^{[x[x(\quad)]]}$ für ganzzahliges Argument (Abundanzen). Mittheilungen der Mathematische Gesellschaft in Hamburg 4 (1901), pp. 33–50. Although modular arithmetic does not work for exponents, here are some striking congruences for hyperpowers.

Meyl [76]. Solution de la question 1196: Résoudre en nombres entiers positifs l'équation $(x+1)^y = x^{y+1} + 1$. Nouvelles Annales de Mathématiques (2) 15 (1876), pp. 545–547.

Mitchelmore, M. C. [74]. A matter of definition. This MONTHLY 81 (1974), pp. 643–647. This is about the iterated exponential.

Moret-Blanc [76]. Solution de la question 1175: Résoudre en nombres entiers et positifs l'équation $x^y = y^x + 1$. Nouvelles Annales de Mathématiques (2) 15 (1876), pp. 44–46.

Moulton, E. J. [16]. The real function defined by $x^y = y^x$. This MONTHLY 23 (1916), pp. 233–237.

Muleškov, A. S. [72]. Endless exponents. (Bulgarian) Godišnik Visš. Tehn. Učebn. Zaved. Mat. 6 (1970), pp. 77–91 (1972). MR48, #2615.

Nederpelt, R. P. [72]. Nesting habits of the laddered parenthesis. This MONTHLY 79 (1972), pp. 395–396.

Nesbitt, A. M. [13]. Discuss the curves (1) $y^x = x^y$, (2) $y = x^{xy}$. Mathematical Questions and Solutions in The Educational Times (2) 23 (1913), pp. 77–79.

Niven, Ivan [72]. Which is larger, e^π or π^e? Two-Year College Math. J. 3 (1972), pp. 13–15.

Oettinger, L. [64]. Ueber den grössten Werth von $\sqrt[x]{x}$ und einige damit zusammenhängende Sätze. Archiv der Mathematik und Physik 42 (1864), pp. 106–114.

Ogilvy, C. S. [49]. Convergence of an infinite involution. Problem E853. This MONTHLY 56 (1949), pp. 555–556.

Pincherle, Salvatore [06]. Funktionaloperationen und Gleichungen. Encyklopädie der mathematischen Wissenschaften mit Einschluss ihrer Anwendungen, vol. 2, 1st part, 2nd half (1906), pp. 761–817, especially pp. 790–797.

Pisanelli, Domingos [52]. Study of a sequence of powers. (Portuguese) Bol. Soc. Mat. São Paulo 6 (1951), pp. 61–68 (1952). MR15, p. 950.

Riordan, John [73]. A note on Catalan parentheses. This MONTHLY 80 (1973), pp. 904–906. MR49, #73.

Rotando, Louis; Plummer, Robert [77]. Can a number be equal to its logarithm? Two-Year College Math. J. 8 (1977), p. 97.

Samelson, Hans [78]. Query #142. Notices of the American Math. Society 25 (1978), pp. 197, 253, 335.

Sato, Daihachiro [72]. Algebraic solution of $x^y = y^x$ $(0 < x < y)$. Proceedings American Math. Society 31 (1972), p. 316. MR44, #5272.

Scheffler, H. [51]. Ueber die durch die Gleichung $y = \sqrt[x]{x}$ dargestellten Kurven. Archiv der Mathematik und Physik 16 (1851), pp. 133–137.

Schimmack, R. [12]. Zur Gleichung $x^y = y^x$. Unterrichtsblätter für Mathematik und Naturwissenschaften 18 (1912), pp. 34–35.

Schinzel, A. [67]. On two theorems of Gelfond and some of their applications. Acta Arith. 13 (1967/68), pp. 177–236. MR36, #5086.

Schlegel, Victor [78]. Lehrbuch der elementaren Mathematik. Wolfenbüttel (Zwissler Verlag) 1878–80. Part 1—Arithmetik und Combinatorik. p. 26. This is about $x^y = y^x$.

Schröder, Ernst [71]. Über iterirte Functionen. Math. Annalen 3 (1871), pp. 296–322.

Schubert, H. [98]. Grundlagen der Arithmetik. Encyklopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen. Teubner, 1898, pp. 26–27. In this are some more references to finite hyperpowers.

Schulze, Emil [77]. Die vierte Rechenstufe. Archiv der Mathematik und Physik 3 (1886), pp. 302–314. The title is another name for hyperpowers.

Schwering, K. [78]. Ueber die Wurzeln der Gleichung $y^x = x^y$. Zeitschrift für Mathematik und Physik 23 (1878), pp. 339–343.

Segal, Irwin H. [75]. Enzyme Kinetics. Wiley-Interscience, 1975. pp. 113–117.

Seidel, Ludwig [73]. Ueber die Grenzwerthe eines unendlichen Potenzausdruckes der Form $x^{x^{x}}$. Abhandlungen der mathematischphysikalischen Classe der Königl. Bayerischen Akademie der Wissenchaften. 11 (1873), pp. 1–10.

Shell, Donald L. [59]. Convergence of Infinite Exponentials. Ph.D. dissertation, University of Cincinnati, 1959.

Shell, Donald L. [62]. On the convergence of infinte exponentials. Proceedings of the American Math. Society 13 (1962), pp. 678–681.

Shilgalis, Thomas W. [73]. Graphical solution of the equation $a^b = b^a$. The Math. Teacher 66 (1973), p. 235.

Sierpiński, W. [58]. Cardinal and Ordinal Numbers, Monografie Matematyczne, vol. 34, Warsaw, 1958. pp. 151, 359–362.

Slobin, H. L. [31]. The solutions of $x^y = y^x$, $x > 0$, $y > 0$, $x \neq y$, and their graphical representation. This MONTHLY 38 (1931), pp. 444–447.

Straffin, Philip D., Jr. [78]. Periodic points of continuous functions. Math. Magazine 51 (1978), pp. 99–105.

Tanturri, A. [15]. Sull'uguaglianza $a^b = b^a$ con a e b interi e positivi. Periodico di Matematica 30 (1915), pp. 186–187.

Thijsen, W. P. [50]. On iterated exponentiation. (Dutch) Simon Stevin 27 (1950), pp. 177–192. MR12, p. 396.

Thron W. J. [57]. Convergence of infinite exponentials with complex elements. Proceedings of the American Math. Society 8 (1957), pp. 1040–1043.

Tijdeman, R. [76]. On the equation of Catalan. Acta Arithmetica 29 (1976), pp. 197–209.

Varner, John T., III [76]. Comparing a^b and b^a using elementary calculus. The Two-Year College Math. J. 7 (1976), p. 46.

Vosper, A. G. [55]. On note 2285: multiple exponentials. Math. Gazette 39 (1955), pp. 141–143.

Wellen, Edward [78]. Bearings. The Magazine of Fantasy and Science Fiction. Feb. 1978, p. 144. The infinite power iteration of e^{-1} even caught the eye of this writer of science fiction.

Wilson, Perry B. [77]. An interesting mathematical sequence and its relation to the threshold for turbulent bunch lengthening. Stanford Linear Accelerator Center (Stanford, California), PEP-232 (Feb. 1977).

Wittstein, Th. [45]. Auflösung der Gleichung $x^y = y^x$ in reellen Zahlen. Archiv der Mathematik und Physik 6 (1845), pp. 154–163.

Wöpcke, F. [51]. Note sur l'expression $\left(((a^a)^a)^{\cdot^{\cdot^{\cdot}}}\right)^a$ et les fonctions inverses correspondantes. Crelle's Journal für die reine und Angewandte Mathematik 42 (1851), pp. 83–90.

Wright, E. M. [47]. Iteration of the exponential function. The Quarterly Journal of Mathematics, Oxford Series 18 (1947), pp. 228–235.

PROOF OF THE FUNDAMENTAL THEOREM OF ALGEBRA

J. L. BRENNER
10 Phillips Road, Palo Alto, CA 94303

R. C. LYNDON
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109

The Fundamental Theorem of Algebra states that every nonconstant polynomial with complex coefficients has a zero in the complex plane. There are many proofs: at least one in every textbook of complex analysis. New ones appear frequently; for examples, and references, see **[1]**, **[5]**, **[6]**, **[7]**, **[9]**, **[10]**, **[11]**. Why, then, propose still another proof?

It is a truism that the theorem is not really a theorem of algebra but of analysis or topology. In the present note we present a proof that ought to be intelligible to a precalculus student; that could be programmed in an introductory course in computing; and that requires no set-theoretic machinery. This last point is significant because nowadays there is not just one set theory: there are several, differing in axioms and content, and surely anything called a fundamental theorem ought to depend as little as possible on set theory.

To isolate the necessary appeal to limits and continuity, and to emphasize how small it is, we do not actually prove the fundamental theorem, but only the following proposition, which contains everything except the brief final stage.

THEOREM. *Let $P(z)$ be a nonconstant polynomial with complex coefficients. Then there is a positive number $S > 0$, depending only on P, with the following property: for every $\delta > 0$ there is a complex number z such that $|z| \leqslant S$ and $|P(z)| < \delta$.*

A similar reduction underlies some of the classical proofs (see, for example, **[5]**, **[7]**, **[9]**), but we emphasize that the proof given here uses only the elementary algebra of the complex numbers and simple inequalities. In essence, it is a version of a recent proof by H. W. Kuhn **[2]**, but modified by ideas of H. Ruderman in Problem 6192 **[8]** and of solvers of this problem, namely O. P. Lossers, Kurt Luoto, and E. S. Rosenthal. We note, in further support of one of our theses, that Kuhn **[3]** has published a usable computer program implementing his ideas.

For the proof, let $P(z) = \sum_{j=0}^{n} a_j z^j$. We suppose, without loss of generality, that $a_0 \neq 0$, and $a_n = 1$. Define $A = \sum_{j=0}^{n} |a_j|$ and $R = 2A$; note that $A > 1$.

LEMMA 1. *If* $|z| \geqslant R$ *then*
(1) $|P(z)| \geqslant A$,
(2) $|P(z) - z^n| \leqslant |z|^n/2$,
(3) $|\operatorname{arc} P(z) - \operatorname{arc}(z^n)| \leqslant \pi/6$,
where arc w *denotes the principal value of the angle of* w.

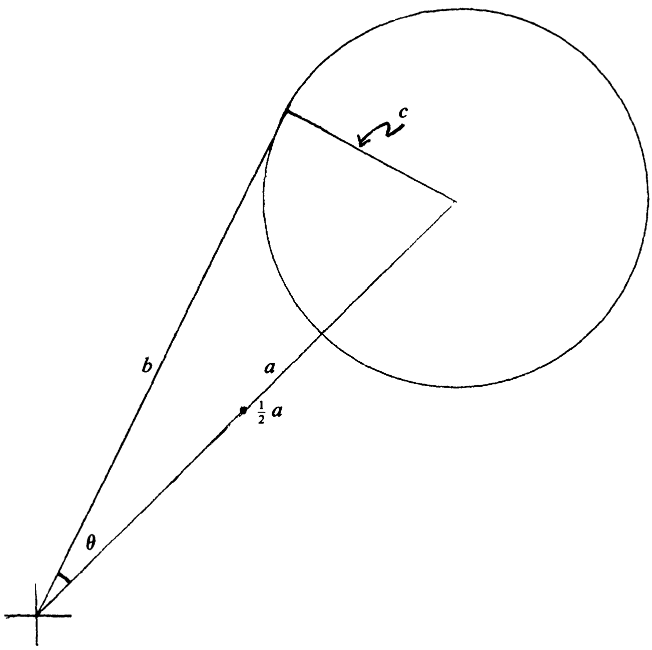
Proof. (1)

$$|P(z)| \geqslant |z^n| - \sum_{0}^{n-1} |a_j||z^j| \geqslant |z^{n-1}|(|z| - A + 1) \geqslant R^{n-1}(R - A) = R^{n-1}A > A.$$

(2) Since $|z| \geqslant 2A - 2$,

$$\begin{aligned}|P(z) - z^n| &\leqslant \sum_{j=0}^{n-1} |a_j||z|^j \leqslant \sum_{j=0}^{n-1} |a_j||z|^{n-1} \\ &= (A-1)|z|^{n-1} = (A-1)|z|^{-1}|z|^n \\ &\leqslant \frac{A-1}{2A-2}|z|^n = |z|^n/2.\end{aligned}$$

The work of the second author was supported by the National Science Foundation under Grant MCS 78-02463.



Since $a = |z^n|$, $b = |P(z)|$, $c \leqslant \frac{1}{2}a$, and since θ is maximal for $b \perp c$, it is clear that $\tan\theta \leqslant \frac{1}{2}$, $\theta \leqslant \pi/6$.

FIG. 1

(3) This follows from (2) by a geometrical argument: see Fig. 1.

We say that the complex number w *lies in the* kth *quadrant* ($k = 1, 2, 3$, or 4) provided $w \neq 0$ and $(k-1)\pi/2 \leqslant \text{arc } w < k\pi/2$. We write $Q(w) = k$. To say that z_1 and z_2 lie in opposite quadrants means that $Q(z_1)$, $Q(z_2)$ differ by 2.

LEMMA 2. *If* z_1, z_2 *lie in opposite quadrants, then* $|z_1|, |z_2| \leqslant |z_2 - z_1|$.

See Fig. 2.

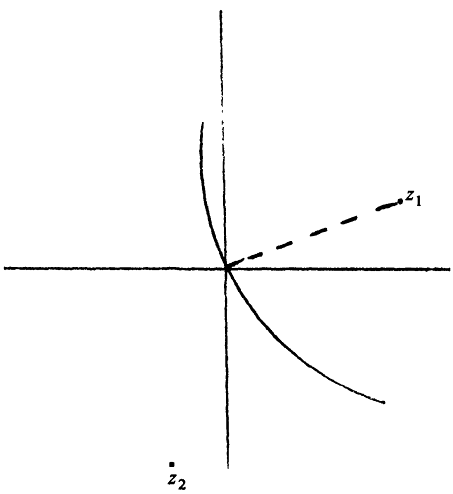


FIG. 2

LEMMA 3. *If $S > 0$, there is a number $k > 0$, depending only on S and P, such that whenever $|z_1|, |z_2| \leqslant S$ we have $|P(z_2) - P(z_1)| \leqslant K|z_2 - z_1|$. (In fact, $K = A\{\max(1, S)\}^n$ is a possible choice for K.)*

Proof. We may suppose $S > 1$. Then

$$P(z_2) - P(z_1) = \sum_{j=0}^{n} a_j(z_2^j - z_1^j)$$

$$= (z_2 - z_1)\sum_{j=1}^{n} a_j \sum_{h+k=j-1} z_1^h z_2^k,$$

$$|P(z_2) - P(z_1)| \leqslant |z_2 - z_1| \sum_{j=1}^{n} |a_j| S^{j-1} \leqslant |z_2 - z_1| S^n A.$$

Now let a given positive number δ be given, $0 < \delta < 1$. We choose an equilateral triangle T enclosing the circle $|z| = R$, and a number $S > 0$ such that T is enclosed in the circle $|z| = S$. Let $\varepsilon = \delta/K$. Let Δ be the closed two-dimensional set inside T. Then, by Lemma 3,

(4) $\qquad z_1, z_2 \in \Delta$ and $|z_1 - z_2| < \varepsilon$ imply that $|P(z_1) - P(z_2)| < \delta$.

We now divide Δ into congruent equilateral triangles Δ_ν by equidistant lines parallel to the sides of T. See Fig. 3. We take the mesh small enough so that each triangle has sides of length no greater than ε and so that, if z and z' are adjacent vertices of the mesh lying on the sides of T, then $|\operatorname{arc} z' - \operatorname{arc} z| < \pi/(6n)$. If z and z' are vertices of the same small triangle Δ_ν, then $|z' - z| < \varepsilon$, whence by (4), $|P(z') - P(z)| < \delta$.

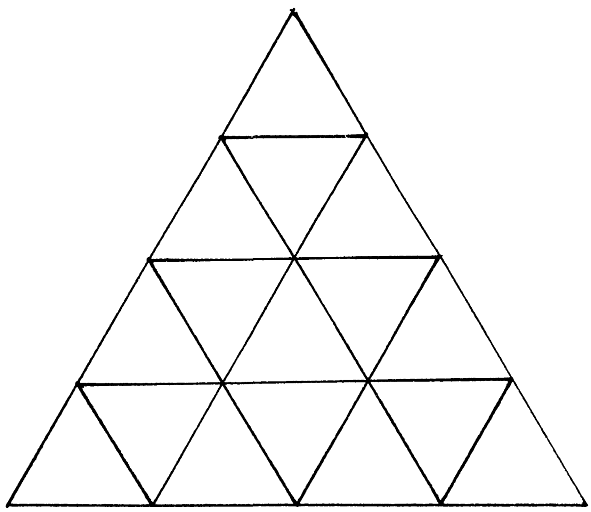


FIG. 3

If some triangle Δ_ν has two vertices z, z' such that $P(z)$, $P(z')$ lie in opposite quadrants, then by Lemma 2 we have $|P(z)| < |P(z') - P(z)| < \delta$, the conclusion of the theorem. To complete the proof we assume that none of the triangles Δ_ν has two vertices z and z' such that $P(z)$ and $P(z')$ lie in opposite quadrants (and no vertex z with $P(z) = 0$), and from this we deduce a contradiction.

Under our assumption, if z and z' are two adjacent vertices of our mesh (that is, two vertices of the same Δ_ν), then $w = P(z)$ and $w' = P(z')$ never lie in opposite quadrants. That is, $Q(w') - Q(w) \equiv d(w, w')$ (modulo 4), where $d(w, w')$ is -1, 0, or $+1$. [$d(w, w')$ is defined by

the preceding sentence.]

Let $z_1,\ldots,z_t$ be the vertices of the mesh that lie on the sides of T, taken in cyclic order in the counterclockwise sense, and let $w_j = P(z_j)$. Then, because we have chosen the mesh small enough, the sequence

$$Q(z_1^n),\ldots,Q(z_t^n), \tag{*}$$

apart from repetitions, runs through the cycle 1, 2, 3, 4 exactly n times. By (3), the sequence

$$Q(w_1),\ldots,Q(w_t) \tag{**}$$

differs from the sequence (*) only in that certain (consecutive) subsequences of the form $h,\ldots,h,\ h',\ldots,h'$, where $h' \equiv h+1$ (modulo 4), are replaced by subsequences of the form $h,\ h_1,\ldots,h_m,\ h'$, where all the h_j are h or h'. Define $D(\Delta) = \Sigma d(w_{j+1},w_j)$, summed over $j=1,\ldots,t$, interpreting $w_{t+1} = w_1$. It follows that $D(\Delta) \equiv \Sigma d(w_{j+1},w_j) = \Sigma d(z_{j+1}^n, z_j^n) = 4n$.

Let z', z'', z''' be the vertices of some triangle Δ_ν, taken in cyclic order counterclockwise. Our hypothesis implies that $Q(P(z'))$, $Q(P(z''))$, and $Q(P(z'''))$ assume at most two adjacent values k and k', where $k' \equiv k+1$ (modulo 4). It follows that $D(\Delta_\nu) = d(w'',w') + d(w''',w'') + d(w',w''') = 0$. Therefore $\Sigma D(\Delta_\nu) = 0$, where the summation is over all triangles Δ_ν of the mesh.

Suppose that z' and z'' are the two ends of an interior edge of the mesh, separating two small triangles Δ_ν and Δ_μ. Then one of the two terms $d(w'',w')$ and $d(w',w'')$ will occur in one of $D(\Delta_\nu)$, $D(\Delta_\mu)$ and the other in the other. Since $d(w',w'') = -d(w'',w')$, these two terms will cancel each other in the sum $\Sigma D(\Delta_\nu)$. It follows that $\Sigma D(\Delta_\nu)$ reduces to the sum of terms $d(w',w'')$ arising from edges on the sides of T, that is, that $\Sigma D(\Delta_\nu) = D(\Delta)$.

Since $D(\Delta) = 4n$, whereas $\Sigma D(\Delta_\nu) = 0$, we have a contradiction.

We thank M. Marden and R. P. Boas for useful criticism.

References

1. C. Fefferman, An easy proof of the fundamental theorem of algebra, this MONTHLY, 74 (1967) 854–855.
2. H. W. Kuhn, A new proof of the fundamental theorem of algebra, Mathematical Programming Study 1, North Holland, 1974, pp. 148–158.
3. ———, Finding roots of polynomials by pivoting, in Fixed Points, Algorithms and Applications, 1977, pp. 12–39.
4. Kurt Luoto, A graph labeling theorem, manuscript.
5. R. E. Moritz, On certain proofs of the fundamental theorem of algebra, this MONTHLY, 10 (1903) 159–160.
6. R. M. Redheffer, What! Another note just on the fundamental theorem of algebra?, this MONTHLY, 71 (1964) 180–185.
7. P. C. Rosenbloom, An elementary constructive proof of the fundamental theorem of algebra, this MONTHLY, 53 (1946) 562–570.
8. H. Ruderman, Problem 6192, this MONTHLY, 86 (1979) 598.
9. Heinrich Weber, Lehrbuch der Algebra, 2nd ed., vol. 1, Braunschweig, 1898, pp. 137–142.
10. S. Wolfenstein, Proof of the fundamental theorem of algebra, this MONTHLY, 74 (1967) 853.
11. H. Zassenhaus, On the fundamental theorem of algebra, this MONTHLY, 74 (1967) 485–497.

MISCELLANEA

53. Who said *that*?

1. "Numerical calculation, and all the devices connected with it, would seem to deserve a far more prominent place in elementary teaching than they receive at present."

2. "The calculating power alone should seem to be the least human of qualities, and to have the smallest amount of reason in it; since a machine can be made to do the work of two or three calculators, and better than any one of them."

(Answers on p. 265.)

THE FIRST PROOF OF THE QUADRATIC RECIPROCITY LAW, REVISITED

EZRA BROWN
Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

1. Introduction and Background. Let p be an odd prime. The integer a, prime to p, is said to be a *quadratic residue* or *nonresidue* of p according as the congruence

$$x^2 \equiv a \pmod{p}$$

is or is not solvable. The Legendre symbol $(a|p)$ is defined to be $+1$ or -1 according as a is a quadratic residue or nonresidue of p.

Because of the relations (to be shown later)

$$(ab|p) = (a|p)(b|p),$$
$$(a^2|p) = (1|p) = 1,$$

and

$$(a|p) = (b|p) \qquad \text{if } a \equiv b \pmod{p},$$

evaluation of the Legendre symbol reduces to evaluating the symbols $(-1|p)$, $(2|p)$, and $(p|q)$, where p and q are distinct odd primes. Because of the obvious difficulties involved in direct evaluation (i.e., calculating the squares mod$p$ for large values of p) of such symbols as

$$(-1|6301), \quad (2|11213) \quad \text{and} \quad (3|6700417),$$

it is desirable to discover some general principles.

The first deep-lying theorem in number theory is a statement of three general principles that are sufficiently powerful to enable one to evaluate any Legendre symbol whatsoever; the theorem is called the Quadratic Reciprocity Law (QRL).

Briefly, the QRL states that, if p and q are distinct odd primes, then

$$(p|q)(q|p) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}; \tag{1}$$

$$(2|p) = (-1)^{(p^2-1)/8}, \tag{2}$$

and

$$(-1|p) = (-1)^{(p-1)/2}. \tag{3}$$

It is now easy to show that

$$(-1|6301) = (-1)^{3150} = 1,$$
$$(2|11213) = (-1)^{15716421} = -1,$$

and

$$(3|6700417) = (6700417|3) = 1 \text{ (exercise!)}.$$

In accordance with tradition, we refer to equation (1) as the QRL and to equations (2) and (3) as the Supplementary Laws.

The elegant symmetry of the QRL was so intriguing that Gauss called it "the gem of the Higher Arithmetic." Gauss gave the first complete proof in 1796—at the advanced age of 19—and subsequently found at least five other proofs. During the nineteenth century, about 50 proofs were published, including a variation on Gauss's first proof by Mathews [7] (see Bachmann [1, Chapter 6] for a list of references, as well as a history); since then, many other

The author received his Ph.D. from Louisiana State University in 1969 under the direction of Gordon Pall. Since then he has been at Virginia Tech, except for a year at the University of Munich. His mathematical interests are in number theory (especially quadratic forms, reciprocity laws, and class-number problems), graph theory, and mathematical applications to the social sciences.—*Editors*

proofs have appeared, so attractive is this theorem. [The title of [4] should be taken with a grain of salt placed firmly on the tongue (-in-cheek).] Note that the QRL says that $(p|q) = (q|p)$ if p or $q \equiv 1 \pmod 4$, and $(p|q) = -(q|p)$ if $p \equiv q \equiv 3 \pmod 4$.

In most textbooks the proof of the QRL is a variation of Gauss's third proof, which is based on the following result, known as Gauss's Lemma:

GAUSS'S LEMMA. *Let p be an odd prime, with $(a,p) = 1$. Reduce the numbers $a, 2a, \dots, (p-1)/2 \cdot a \pmod p$ to the interval $\{1, \dots, p-1\}$. Let u denote the number of these reduced residues that lie between $p/2$ and p. Then*

$$(a|p) = (-1)^u.$$

Gauss then showed that

$$\sum_{j=1}^{(p-1)/2} [qj/p] + \sum_{j=1}^{(q-1)/2} [pj/q] = (p-1)(q-1)/4, \tag{*}$$

the QRL following from (*) and Gauss's Lemma.

However, the textbook proof customarily ends with a clever lattice-point-counting proof of (*) due to Eisenstein, a student of Gauss and considered by the latter to be one of the three greatest mathematicians of all time—along with Newton and Archimedes. [Gauss was modest.]

Why is the first proof never mentioned?

Gauss wrote that he discovered the QRL experimentally in March of 1795; that he spent many sleepless nights searching for a proof; and that his search was finally rewarded on April 8, 1796. This proof was published in 1801 as Articles 125-145 in his monumental treatise on number theory, *Disquisitiones Arithmeticae* [3], and was found by induction. It was not a masterpiece of exposition: as the English number theorist Henry John Stephen Smith wrote [8], this first proof was "presented by Gauss in a form very repulsive to any but the most laborious students."

Not for the faint-hearted, that first proof!

However, a number of years ago, Gordon Pall succeeded in making the first proof readable by an ingenious device: he recast it as a proof by descent. Together with some other modifications of our own (made while digging through the original Latin version of the *Disquisitiones*), it is the most direct proof, and hence clearly shows just exactly why the QRL is true. It also avoids Gauss's Lemma (as do many other proofs—see, for example, [2]).

It is our purpose here to present this proof.

An historical note: the Supplementary Laws were known to Fermat and Euler, respectively. Euler found proofs of these, determined the value of $(a|p)$ for a few small values of a, knew the Product Rule $(ab|p) = (a|p)(b|p)$, and stated an equivalent form of the QRL in 1783, without proof. Lagrange "came close" in 1775, but never actually stated the Law. Legendre guessed the QRL in 1785 and published a "proof" in the same memoir (see [6]); unfortunately, his reasoning was insufficient—as we shall see.

2. A Few Lemmas. The proof is based on a sequence of lemmas. Throughout, lower-case Latin letters stand for integers, p and q stand for primes, $a|b$ means that a is a divisor of b and (a,b) is the greatest common divisor of a and b.

LEMMA 1. (a) *If $(a,p) = 1$ and $x^2 \equiv a \pmod p$ has a solution x, then the only solutions of this congruence are x and $-x \pmod p$.*

(b) $(a^2|p) = (1|p) = 1$; *and if $a \equiv b \pmod p$, then $(a|p) = (b|p)$.*

Proof. (a) Note that $x^2 \equiv y^2 \pmod p$ implies $p|(x-y)(x+y)$; but p is a prime, so either $p|x-y$ or $p|x+y$. Hence $x \equiv y$ or $x \equiv -y \pmod p$.

(b) These are direct consequences of the definition of the Legendre symbol.

Proof. Let $1 \leqslant x \leqslant p-1$; by the a-inverse of x we mean the unique solution of $yx \equiv a \pmod p$. Note that x and its a-inverse are incongruent mod$p$ unless $r^2 \equiv a \pmod p$ is solvable. If the latter is true, then r is its own a-inverse, as is $-r$, and, by Lemma 1, $\pm r$ are the only self-a-inverses (mod$p$). As p is odd, $r \not\equiv -r \pmod p$.

If $(a|p) = -1$, then the numbers $1,\ldots,p-1$ fall into $(p-1)/2$ distinct pairs of a-inverses, each with product $a \pmod p$. Hence, if $(a|p) = -1$, then

$$\begin{aligned} a^{(p-1)/2} &\equiv 1\cdot 2\cdots(p-1) \pmod p \\ &\equiv -(a|p)(p-1)! \pmod p. \end{aligned}$$

If $(a|p) = 1$, then the numbers $1,\ldots,p-1$ fall into $((p-1)/2)-1$ distinct pairs of a-inverses, each with product $a \pmod p$, together with the pair of self-a-inverses $\pm r \pmod p$, where $r^2 \equiv a \pmod p$. Since $r(-r) \equiv -a \pmod p$, we have

$$a^{((p-1)/2)-1}(-a) \equiv 1\cdot 2\cdots(p-1) \pmod p,$$

so that

$$\begin{aligned} a^{(p-1)/2} &\equiv -1\cdot 2\cdots(p-1) \pmod p \\ &\equiv -(a|p)(p-1)! \pmod p. \end{aligned}$$

From this lemma we deduce, with surprising ease, a number of useful and famous corollaries.

LEMMA 3. *If* $(a,p) = (b,p) = 1$, *then*
(a) $-1 \equiv (p-1)! \pmod p$ (*Wilson's Theorem*);
(b) $a^{(p-1)/2} \equiv (a|p) \pmod p$ (*Euler's Theorem*);
(c) $a^{p-1} \equiv 1 \pmod p$ (*Fermat's Little Theorem*);
(d) $(-1|p) = (-1)^{(p-1)/2}$ (*Supplement to QRL*);
(e) $(a|p)(b|p) = (ab|p)$ (*Product Rule for the Legendre symbol*).

Proof. (a) follows by setting $a = 1$ in Lemma 2 and from the fact that $(1|p) = 1$.
(b) follows from (a) and the substitution of -1 for $(p-1)!$ on the right side of Lemma 2.
(c) follows by squaring both sides in (b).
(d) follows from (b) with $a = -1$ and from the fact that $1 \equiv -1 \pmod p$ is impossible for odd primes p.
(e) follows from (b):

$$\begin{aligned} (a|p)(b|p) &\equiv a^{(p-1)/2}b^{(p-1)/2} \pmod p \\ &\equiv (ab)^{(p-1)/2} \pmod p \\ &\equiv (ab|p) \pmod p, \end{aligned}$$

and

$$1 \not\equiv -1 \pmod p,$$

so

$$(a|p)(b|p) = (ab|p).$$

3. The Quadratic Character of 2. We may now proceed to the supplementary law for the prime 2. The proof is partly by descent, follows the original treatment by Gauss (Articles 212–214 of the *Disquisitiones*), and is independent of Gauss's Lemma.

THE SUPPLEMENT TO THE QRL FOR THE PRIME 2. *If p is an odd prime, then*

$$(2|p) = (-1)^{(p^2-1)/8}.$$

That is, $(2|p)=1$ *if* $p\equiv 1$ *or* $7 \pmod 8$, *and* $(2|p)=-1$ *if* $p\equiv 3$ *or* $5 \pmod 8$.

Proof. *Case* $p\equiv 3$ or $5 \pmod 8$. If the theorem is false, then there is a least positive prime $p\equiv 3$ or $5 \pmod 8$ for which $(2|p)=1$. Then there exists x such that $1\leqslant x\leqslant p-1$ and

$$x^2\equiv 2 \pmod p.$$

We may assume that x is odd (just replace x by $p-x$ if not), so that $x^2=2+pm$ for some integer m; hence

$$1\equiv x^2\equiv 2\pm 3m \pmod 8,$$

and

$$m\equiv \pm 5 \pmod 8.$$

Now a product of primes $8n\pm 1$ is again of that form; so m has a prime divisor $q\equiv \pm 3 \pmod 8$. Moreover,

$$1\leqslant 2+pm=x^2<p^2,$$

so that $q\leqslant m<p$. But $q|m$, so that

$$x^2=2+pm\equiv 2 \pmod q, \qquad (2|q)=1,$$

and this contradicts the minimality of p.

Case $p\equiv 7 \pmod 8$. Using the method of descent, as in the previous case, we similarly show that, if $p\equiv 5$ or $7 \pmod 8$, then $(-2|p)=-1$. Thus, if $p\equiv 7 \pmod 8$, then $(2|p)=(-1|p)(-2|p)=(-1)(-1)=1$ by Lemma 3 (d) and (e).

Case $p\equiv 1 \pmod 8$. First, the numbers $1^2,2^2,\dots,((p-1)/2)^2$ are all distinct $(\bmod p)$, and $x^2\equiv(-x)^2 \pmod p$. Hence there are exactly $(p-1)/2$ quadratic residues, and $(p-1)/2$ quadratic nonresidues, mod p. Let s be a quadratic nonresidue of $p=8n+1$. By Euler's Theorem,

$$-1\equiv s^{(p-1)/2}\equiv s^{4n} \pmod p,$$

so that

$$s^{2n}+1\not\equiv 0,\quad s^{4n}+1\equiv 0 \pmod p.$$

Hence,

$$\begin{aligned}
2&\equiv 2s^{2n}s^{-2n} && \pmod p\\
&\equiv s^{-2n}(2s^{2n}+s^{4n}+1) && \pmod p\\
&\equiv s^{-2n}(s^{2n}+1)^2 && \pmod p\\
&\equiv \left(s^{-n}(s^{2n}+1)\right)^2 && \pmod p\\
&\not\equiv 0 && \pmod p.
\end{aligned}$$

Hence, $(2|p)=1$. (Challenge Problem: find a proof that $(2|p)=1$ for $p\equiv 1 \pmod 8$ *by descent*!)

4. Two More Lemmas. Before we prove the QRL, we find it useful to introduce a generalization of the Legendre symbol. Let us define the *Jacobi symbol* $(a|b)$ for integers a and b with $(a,b)=1$ and b odd by the equations

$$(a|b)=(a|-b), \qquad (a|1)=1, \qquad (a|b)=\prod_{i=1}^{n}(a|p_i),$$

where $b=p_1\cdots p_n$ is a product of primes and $(a|p_i)$ is the Legendre symbol.

Let us further say that, if $(a,b)=1$, then the two odd integers a and b are *in tune* if

$$(a|b)(b|a)=(-1)^{\frac{a-1}{2}\cdot\frac{b-1}{2}}$$

and *out of tune* otherwise. By Lemma 3(d), -1 and p are in tune for any odd prime p, whereas

-1 and $-p$ are out of tune.

LEMMA 4. *Let a, b, c, and d be integers, with $(2a, bc) = (d, b) = 1$. Then*:

(a) $(a|b)(a|c) = (a|bc)$ *and* $(a|b)(d|b) = (ad|b)$.

(b) *If* $x^2 \equiv a \pmod b$ *has a solution, then* $(a|b) = 1$.

(c) *If, in addition, a is odd, and if two of the three pairs* $\{a, b\}$, $\{a, c\}$, *and* $\{a, bc\}$ *are in tune, so is the third (the 2-out-of-3 Rule).*

Proof. (a) follows from the definition and the Product Rule for the Legendre symbol.

(b) If $x^2 \equiv a \pmod b$ has a solution, so does $x^2 \equiv a \pmod{p_i}$ for each i, where $b = p_1 \cdots p_n$ is a product of primes. Hence, by the definition we have

$$(a|b) = \prod_{i=1}^{n} (a|p_i) = \prod_{i=1}^{n} 1 = 1.$$

(c) First note that, by (a),

$$\begin{aligned}(a|b)(b|a)(a|c)(c|a)(a|bc)(bc|a) &= (a|b)(a|c)(a|bc)(b|a)(c|a)(bc|a)\\ &= (a|bc)^2(bc|a)^2\\ &= 1,\end{aligned}$$

so that the product of the left sides of all three "in-tune" statements is 1. Then observe that

$$\frac{b-1}{2} + \frac{c-1}{2} + \frac{bc-1}{2} = \frac{(b+1)(c+1)-4}{2};$$

this is an even integer for all odd b and c, so that the exponent

$$\frac{a-1}{2}\left(\frac{b-1}{2} + \frac{c-1}{2} + \frac{bc-1}{2}\right)$$

is even. From this we deduce that the product of the right sides of all three "in tune" statements is 1. Hence, it is impossible that two of the statements be true and the third false, and the 2-out-of-3 Rule follows.

As mentioned before, Legendre came close to proving the QRL, but he used the following lemma without sufficient proof.

LEMMA 5. *If q is a prime of the form $4n + 1$, then there exists a prime $p < q$ such that* $(q|p) = -1$.

Proof. (a) If $q \equiv 5 \pmod 8$ then $q - 2 \equiv 3 \pmod 8$, so that $q - 2$ has a prime factor $p \equiv 3$ or $5 \pmod 8$. Hence, $q \equiv 2 \pmod p$ and so

$$(q|p) = (2|p) = -1$$

by the Supplementary QRL for 2.

(b) If $q \equiv 1 \pmod 8$ and the lemma is false, then

$$x^2 \equiv q \pmod p$$

has a solution for each odd prime $p < q$, as does

$$x^2 \equiv q \pmod 8 \qquad (\text{let } x = 1!).$$

By a standard procedure (see [**5**, p. 98]), it follows that

$$x^2 \equiv q \pmod{p^r}$$

has a solution for all $r \geqslant 1$ (respectively, $r \geqslant 3$) if p is odd (respectively, $p = 2$). Hence, by the Chinese Remainder Theorem, we may solve the congruence

$$x^2 \equiv q \pmod{(2k+1)!}$$

where k satisfies

$$k^2 < q < (k+1)^2$$

and hence

$$2k + 1 < 2\sqrt{q} + 1 < q.$$

We proceed to derive a contradiction from this.

First, $((2k+1)!, q) = 1$, so that $((2k+1)!, x) = 1$. Next,

$$(x+k)\cdots x \cdots (x-k) \equiv 0 \ (\mathrm{mod}\ (2k+1)!)$$

because the binomial coefficient

$$\binom{x+k}{2k+1} = \frac{(x+k)\cdots(x+k-(2k+1)+1)}{(2k+1)!}$$

is an integer. Hence, using $((2k+1)!, x) = 1$, we have

$$\begin{aligned} 0 &\equiv x(x+k)\cdots(x+1)(x-1)\cdots(x-k) && (\mathrm{mod}\ (2k+1)!) \\ &\equiv (x^2-k^2)\cdots(x^2-1^2) && (\mathrm{mod}\ (2k+1)!) \\ &\equiv (q-k^2)\cdots(q-1) && (\mathrm{mod}\ (2k+1)!) \end{aligned}$$

Finally,

$$\begin{aligned} (2k+1)! &= (k+1+k)\cdots(k+1)\cdots(k+1-k) \\ &= ((k+1)^2-k^2)\cdots(q-1^2)(k+1) \\ &> (q-k^2)\cdots(q-1^2)\cdot 1, \end{aligned}$$

so that we have a nonzero integer divisible by a larger integer! This is our desired contradiction, and Lemma 5 follows.

5. The Quadratic Reciprocity Law, by Descent. We are finally ready to prove the QRL, which we may restate as follows:

QUADRATIC RECIPROCITY LAW. *Any two distinct odd primes are in tune.*

Proof of the QRL by descent. If the QRL is false, then there is a least prime q for which there is a smaller prime p such that p and q are out of tune. By the minimality of q, p is in tune with all odd primes less than q; hence, by the 2-out-of-3 Rule and induction, p is in tune with all integers each of whose prime factors are less than q. The proof divides naturally into three cases.

Case $(p|q) = 1$. We can solve $x^2 \equiv p \ (\mathrm{mod}\, q)$; as in our study of the quadratic character of 2, we may assume that x is even and $|x| < q$. Thus

$$x^2 = p + qs$$

with s odd and $|s| < q$, the latter following by

$$|qs| = |x^2 - p| < x^2 < q^2, \qquad |s| < |q| = q.$$

Hence, p is in tune with s.

If $(p, s) = 1$, then

$$x^2 \equiv p \ (\mathrm{mod}\, qs) \quad \text{and} \quad x^2 \equiv qs \ (\mathrm{mod}\, p)$$

imply that

$$(p|qs) = (qs|p) = 1.$$

Because x is even, we have $0 \equiv p + qs \ (\mathrm{mod}\, 4)$, so that not both p and qs are of the form $4n+3$. Hence, p and qs are in tune, so that by the 2-out-of-3 Rule, p and q are in tune—a contradiction.

If $p|s$, set $x = px_1$ and $s = ps_1$; then

$$px_1^2 = 1 + qs_1.$$

Hence,

$$(p|-qs_1) = 1 = (-qs_1|p) \quad \text{and} \quad 1 \equiv -qs_1 \pmod 4,$$

so that p and $-qs_1$ are in tune, as are p and $-s_1$ (for, $|s_1| < q$). We deduce, as before, that p and q are in tune—again, a contradiction.

Case $(p|q) = -1$, $q \equiv 3 \pmod 4$. As $(-1|q) = -1$, we have $(-p|q) = 1$ so that we may write

$$x^2 = -p + qs$$

with x even, s odd, and $0 < s < q$.

If $(p,s) = 1$, we find, as before, that $(qs|-p) = (-p|qs) = 1$, with $-p$ or $qs \equiv 1 \pmod 4$. Hence, $-p$ and qs are in tune, as are -1 and qs. Thus, p and qs are in tune and p and s are in tune (as $|s| < q$), so p and q are in tune.

If $p|s$, we put $s = ps_1$, $x = px_1$, so that

$$1 = -px_1^2 + qs_1.$$

Then $(qs_1|-p) = 1 = (-p|qs_1)$, $-p$ and qs_1 are in tune, p and qs_1 are in tune (2-out-of-3), p and s_1 are in tune ($|s_1| < q$), p and q are in tune.

Case $(p|q) = -1$, $q \equiv 1 \pmod 4$. This is the most difficult case, for we cannot directly write down a quadratic congruence as we could in the previous cases. So we enlist the aid of Lemma 5, Legendre's Waterloo.

Let us choose, by Lemma 5, an odd prime $p_1 < q$ for which $(q|p_1) = -1$. If $(p_1|q) = 1$, then we proceed as in the first case to obtain a contradiction; so let us assume that $(p_1|q) = -1$. As $q \equiv 1 \pmod 4$, we know that p_1 is in tune with q; moreover,

$$(pp_1|q) = (p|q)(p_1|q) = (-1)(-1) = 1$$

by the Product Rule. Hence we may write

(a) $x^2 = pp_1 + qs$	if $(pp_1,s) = 1$;
(b) $pp_1x^2 = 1 + qs_1$	if $(pp_1,s) = pp_1$;
(c) $px^2 = p_1 + qs_2$	if $(pp_1,s) = p$;
(d) $p_1x^2 = p + qs_3$	if $(pp_1,s) = p_1$.

As before we may assume x even and $|s| < q$, so that $|s_i| < q$. The details of the four cases are similar; so we examine only case (c) more closely.

We have

$$(qs_2|p) = (-p_1|p), \quad (qs_2|p_1) = (p|-p_1), \quad \text{and} \quad (p|qs_2) = (p_1|qs_2).$$

Hence,

$$(pp_1|qs_2) = 1 \quad \text{and} \quad (qs_2|pp_1) = (-p_1|p)(p|-p_1).$$

If $p \equiv 3$ and $p_1 \equiv 1 \pmod 4$, then $qs_2 \equiv -p_1 \equiv pp_1 \equiv 3 \pmod 4$, so that

$$(pp_1|qs_2)(qs_2|pp_1) = -(p_1|p)(p|p_1) = -1,$$

since $p_1 \not< q$ implies that p is in tune with p_1. Hence pp_1 is in tune with qs_2, and, as above, this means that p and q are in tune.

If $p \equiv 1$ or $p_1 \equiv 3 \pmod 4$, then either $-p_1$ or $pp_1 \equiv 1 \pmod 4$. As before, we deduce that pp_1 is in tune with qs_1, so that p is in tune with q.

The Quadratic Reciprocity Law follows!

References

1. P. Bachmann, Niedere Zahlentheorie (Part 1), Teubner, Leipzig, 1902.

2. J. S. Frame, A short proof of the quadratic reciprocity law, this MONTHLY, 85 (1978) 818–819.

3. C. F. Gauss, Disquisitiones Arithmeticae, Fleischer, Leipzig, 1801; reprinted in vol. 1 of Gauss's Werke: Disq. Arith.

4. M. Gerstenhaber, The 152nd proof of the law of quadratic reciprocity, this MONTHLY, 70 (1963) 397–398.
5. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford, 1960.
6. A. M. Legendre, Essai sur la théorie des nombres, Paris, 1798.
7. G. B. Mathews, Theory of Numbers, Cambridge, 1892.
8. H. J. S. Smith, Report on the theory of numbers, Reports of the British Association, 1859–1895; reprinted by Chelsea, 1968.

AN ELEMENTARY PROOF OF THE NO-RETRACTION THEOREM

YAKAR KANNAI
Department of Theoretical Mathematics, Weizmann Institute of Science, Rehovot, Israel

1. The no-retraction theorem [**1**], [**6**], [**7**, p. 271] asserts the nonexistence of a continuous function f mapping the unit ball $B \subset R^n$ into its boundary S, such that $f(x) = x$ for all $x \in S$. This theorem is usually proved by means of either combinatorial arguments, homology theory, differential forms, or methods from differential topology, see [**1**], [**5**], [**6**], [**10**]. The proof given in [**3**], while analytic and entirely elementary, uses a homotopy in order to get the desired contradiction via $(n + 1)$-dimensional integration. We offer here a self-contained proof, inspired

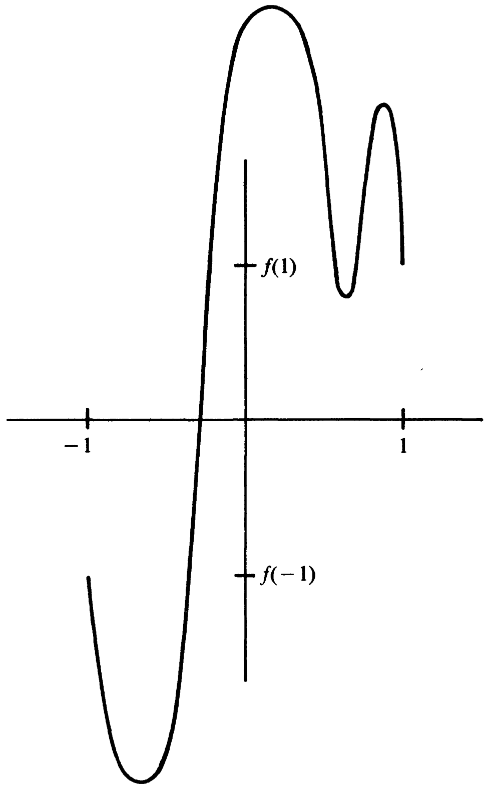


FIG. 1

Yakar Kannai received his Ph.D. in 1968 under S. Agmon at the Hebrew University of Jerusalem. Since 1974 he has held a position at the Weizmann Institute of Science, Rehovot, Israel. He also held visiting positions in a number of American universities and at the Institute for Advanced Study, Princeton, N.J. His research interests lie in the theory of linear partial differential equations, in game theory, and in mathematical economics.

by the one in [3], of the "no differentiable retraction" theorem, a proof which employs only "engineering type" Advanced Calculus concepts. No homotopy is involved in our argument.

To motivate our method, consider the following *one*-dimensional no differentiable retraction theorem:

There exists no differentiable $f:[-1,1]\to\{-1,1\}$, *such that* $f(1)=1, f(-1)=-1$.

(Of course, you do not need the differentiability assumption, as this theorem is just the intermediate value theorem. But having no higher dimensional analogue of the latter theorem—other than the no-retraction theorem itself—we prefer to suggest a proof which we will then generalize to n dimensions.)

Proof. Clearly, $f'(t)\equiv 0$ for $-1\leqslant t\leqslant 1$, since otherwise the range of f would contain an interval. But

$$0=\int_{-1}^{1} f'(t)\,dt=f(1)-f(-1)=1-(-1)=2\neq 0, \tag{1}$$

a contradiction.

In the n-dimensional case, it is only the evaluation of the integral which is slightly more complicated; the idea is that we replace f' by the Jacobian determinant in the evaluation of the (signed) measure of the image (of B under f) and that this measure is determined only by the behavior of f at the boundary S. As we see in Fig. 1, the (signed) measure of the image of an interval $[a,b]$ under the map f is equal to $f(b)-f(a)$; as the interval $[f(a),f(b)]$ is "assumed" once, the other regions and the other times this interval gets "covered" cancel each other out. (In Fig. 1 we exhibit the special case $a=-1$, $b=1$, $f(a)=a$, $f(b)=b$.) The same is true in n variables: The signed measure of the image of B under a map f, such that f is a diffeomorphism on S, is equal to the measure of the set bounded by $f(S)$ (i.e., the bounded component of R^n whose boundary is $f(S)$)—other regions canceling each other out. (Degree theory makes this idea precise, but our integration argument avoids any counting.) The fundamental Theorem of Calculus will be replaced in n-dimensions by Gauss's theorem—see, e.g., [2], where it is argued that Gauss's theorem could be regarded as a proper analogue of the fundamental theorem of Calculus.

In Section 2 we will carry out the proof in n-dimensions. At the end of Section 2 we exhibit a condensed version of our proof, using the theory of differential forms. In Section 3 we outline the well-known derivation of the Brouwer fixed-point theorem from the no-retraction theorem. An analytic proof of the Brouwer fixed-point theorem has appeared recently [8]. We feel that our proof is more motivated.

I am very much indebted to Professor H. Scarf for encouraging discussion concerning this work.

2. We prove in the present section the following:

"NO DIFFERENTIABLE RETRACTION" THEOREM. *There exists no twice differentiable map f of the unit ball B in R^n into its boundary S, such that $f(x)=x$ for all $x\in S$.*

Proof. Let f be such a retraction, $f(x)=(f_1(x),\dots,f_n(x))$. Let $J(x)$ denote the Jacobian determinant of f at x. Expanding $J(x)$ by the first column, we get

$$J(x)=\sum_{i=1}^{n}(-1)^{i+1}\frac{\partial f_1}{\partial x_i}E_i(x) \tag{2}$$

Answers to the questions on p. 256.

1. A. Weil, this MONTHLY, 61 (1954) 36.
2. Oliver Wendell Holmes, *The Autocrat of the Breakfast Table*, 1857, Part 1.

where $E_i(x)$ is the determinant of the matrix obtained from the matrix

$$M(x) = \begin{bmatrix} \dfrac{\partial f_2}{\partial x_1}, \dots, \dfrac{\partial f_n}{\partial x_1} \\ \vdots \\ \dfrac{\partial f_2}{\partial x_n}, \dots, \dfrac{\partial f_n}{\partial x_n} \end{bmatrix} \tag{3}$$

by omitting the ith row. Note that $J(x)$ vanishes identically on B, as the n scalar functions $f_1, \dots, f_n$ satisfy the functional relation $\sum_{i=1}^n f_i^2(x) \equiv 1$. (Note that we use here only the easy part of the vanishing Jacobian theorem.) Integrating $J(x)$ over B, we find, using the rule for differentiation of a product and (2), that

$$\begin{aligned} 0 &= \int_B \cdots \int J(x)\, dx_1 \cdots dx_n \\ &= \int_B \cdots \int \sum_{i=1}^n (-1)^{i+1} \frac{\partial}{\partial x_i}(f_1 E_i)\, dx_1 \cdots dx_n + \int_B \cdots \int \sum_{i=1}^n (-1)^i f_1 \frac{\partial E_i}{\partial x_i} dx_1 \cdots dx_n. \end{aligned} \tag{4}$$

According to a well-known theorem of Jacobi **[4]**, **[9]** (used in the proof of the Brouwer fixed-point theorem in **[3]**),

$$\sum_{i=1}^n (-1)^i \frac{\partial E_i}{\partial x_i}(x) \equiv 0. \tag{5}$$

(If $n = 2$, then (5) reduces to the equality of the mixed derivatives.) To prove (5), let $c_{i,j}(x)$, $i \neq j$, denote the determinant of the matrix obtained from $M(x)$ by omitting the ith row and replacing the row

$$\left(\frac{\partial f_2}{\partial x_j}, \dots, \frac{\partial f_n}{\partial x_j} \right) \text{ by } \left(\frac{\partial^2 f_2}{\partial x_i \partial x_j}, \dots, \frac{\partial^2 f_n}{\partial x_i \partial x_j} \right).$$

Applying the rule for differentiating determinants we see that $\partial E_i / \partial x_i = \sum_{j \neq i} c_{i,j}$. The equality of the mixed derivatives $\partial^2 f_k / \partial x_i \partial x_j = \partial^2 f_k / \partial x_j \partial x_i$ implies that $c_{j,i} = (-1)^{j-i-1} c_{i,j}$ as the row of the second-order derivatives get shifted $j - i - 1$ rows when one passes from $c_{i,j}$ to $c_{j,i}$ if $i < j$, and $i - j - 1$ rows otherwise. Hence

$$\begin{aligned} \sum_{i=1}^n (-1)^i \frac{\partial E_i}{\partial x_i} &= \sum_{i=1}^n (-1)^i \left[\sum_{j<i} c_{i,j} + \sum_{j>i} c_{i,j} \right] \\ &= \sum_{j<i} (-1)^i c_{i,j} + \sum_{j>i} (-1)^i (-1)^{j-i-1} c_{j,i} = 0. \end{aligned}$$

Substituting (5) in (4), we find that a contradiction would follow once we prove that

$$I = \int_B \cdots \int \sum_{i=1}^n (-1)^{i+1} \frac{\partial}{\partial x_i}(f_1 E_i)\, dx_1 \cdots dx_n \neq 0. \tag{6}$$

We transform I into a surface integral according to Gauss's divergence theorem **[2]**, **[11]**, applied to the vector field whose ith component is $(-1)^{i+1} f_1(x) E_i(x)$. We denote by $d\sigma$ the surface element ($(n-1)$-dimensional volume) on the unit sphere S, and utilize the fact that the outward unit normal of S coincides with $x = (x_1, \dots, x_n)$. Hence

$$I = \int_S \cdots \int f_1(x) \sum_{i=1}^n (-1)^{i+1} x_i E_i(x)\, d\sigma. \tag{7}$$

In order to calculate I, observe that $f_i(x) \equiv x_i$ on S, $1 \leqslant i \leqslant n$. Hence $\operatorname{grad} f_i - \operatorname{grad} x_i$ is perpendicular to S there. Thus there exist scalars λ_i (depending on x) such that $\operatorname{grad} f_i(x) =$

$\operatorname{grad} x_i + \lambda_i x$, and the matrix M can be written as

$$\begin{bmatrix} \lambda_2 x_1 & ,\dots, & \lambda_n x_1 \\ 1+\lambda_2 x_2 & & \\ \vdots & & \vdots \\ \lambda_2 x_n & & 1+\lambda_n x_n \end{bmatrix}.$$

The sum $\Sigma_{i=1}^{n}(-1)^{i+1}x_i E_i(x)$ is equal to the determinant

$$\begin{vmatrix} x_1 & \lambda_2 x_1 & \cdots & \lambda_n x_1 \\ x_2 & 1+\lambda_2 x_2 & & \lambda_n x_2 \\ \vdots & & & \vdots \\ x_n & \lambda_2 x_n & & 1+\lambda_n x_n \end{vmatrix} = \begin{vmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & 1 & & 0 \\ \vdots & & & \vdots \\ x_n & 0 & & 1 \end{vmatrix} = x_1.$$

Moreover, $f_1(x) = x_1$ on S. Inserting these results in (7), we get the result $I = \int_S \cdots \int x_1^2 d\sigma > 0$, contradicting (4). (It is easy to compute that $I = (1/n)\int_S \cdots \int d\sigma = \operatorname{Vol}(B)$.)

Note that the passage from the volume integral (4) to the surface integral (7) corresponds to the two left equalities in (1) and that the evaluation of I as being equal to $\operatorname{Vol}(B)$ corresponds to the fact that the one-dimensional volume of the interval $[-1,1]$ is equal to 2, the right-hand side of (1).

REMARK 1. Only a slight modification of the argument is needed to prove the nonexistence of a differentiable retraction for a general bounded open subset B of R^n with a smooth boundary S, provided that the divergence theorem holds for B and S **[2]**, **[11]**. In fact, all the arguments leading to the formula (7) are valid (with no change) in the general case too. Let $\gamma = (\gamma_1, \dots, \gamma_n)$ denote the outward unit normal of S, and let $d\sigma$ denote the surface element on S. The divergence theorem implies that

$$I = \int_S \cdots \int f_1(x) \sum_{i=1}^{n} (-1)^{i+1}\gamma_i E_i(x)\, d\sigma. \tag{7$'$}$$

Observe that $\operatorname{grad} f_i(x) = \operatorname{grad} x_i + \lambda_i \gamma$ on S, $i = 1, \dots, n$. Hence I can be written in the form (compare the argument following (7))

$$I = \int_S \cdots \int f_1(x)\gamma_1(x)\, d\sigma = \int_S \cdots \int x_1 \gamma_1\, d\sigma. \tag{8}$$

Applying the divergence theorem once more (to the vector $(x_1, 0, \dots, 0)$) we have

$$\operatorname{Vol}(B) = \int_B \cdots \int 1\, dx_1 \cdots dx_n = \int_B \cdots \int \frac{\partial x_1}{\partial x_1} dx_1 \cdots dx_n = \int_S \cdots \int x_1 \gamma_1\, d\sigma.$$

REMARK 2. The proof presented in this note can be cast in a compact form using concepts and results from the theory of differential forms. In fact, define two $(n-1)$-forms on B by

$$\omega_1 = f_1\, df_2 \wedge \cdots \wedge df_n, \qquad \omega_2 = x_1\, dx_2 \wedge \cdots \wedge dx_n.$$

Then

$$d\omega_2 = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n, \qquad d\omega_1 = df_1 \wedge df_2 \wedge \cdots \wedge df_n = 0$$

(as the differentials $df_1, \dots, df_n$ are linearly dependent). Note also that the restrictions to S of ω_1 and ω_2 coincide, as $f_i \equiv x_i$ on S, $1 \leqslant i \leqslant n$. By Stokes's theorem,

$$0 = \int_B d\omega_1 = \int_S \omega_1 = \int_S \omega_2 = \int_B d\omega_2 = \operatorname{Vol}(B) \neq 0. \tag{9}$$

Note the similarity between (9) and (1).

3. The Brouwer fixed-point theorem follows from the no differentiable retraction theorem in a well-known way (see, e.g., [3]). We sketch the argument for completeness. Suppose that $g: B \to B$ is a fixed-point-free continuous map. The compactness of B implies that $|g(x) - x| \geqslant \varepsilon > 0$ for $x \in B$. Let $h(x)$ be a C^2 function such that $|h(x) - g(x)| < \varepsilon/2$ on B and such that $h: B \to B$ (we can even let h be a polynomial). Then $h(x) \neq x$ for $x \in B$, and let $f(x)$ denote (for $x \in B$, $X \notin S$) the unique point on S such that $h(x)$, x and $f(x)$ lie on the same line and x is between $h(x)$ and $f(x)$ (see Fig. 2), $f(x) = x$ for $x \in S$. Then $f(x)$ is a C^2 retraction,

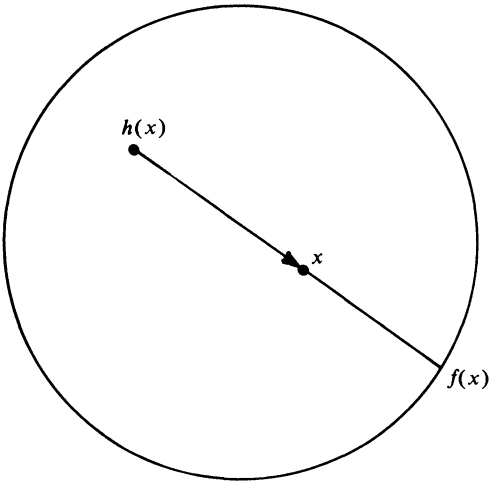


FIG. 2

contradicting the theorem of Section 2. It is not difficult to show (using a suitable approximation argument) that the no differentiable retraction theorem implies directly that there exists no continuous retraction. Using Remark 1 it follows, e.g., that there exists no continuous retraction of the torus onto its boundary.

The research described in this paper was partially supported by NSF and ONR grants while the author was visiting the Cowles Foundation at Yale University.

References

1. W. M. Boothby, On two classical theorems of algebraic topology, this MONTHLY, 78 (1971) 237–249.
2. R. Courant, Differential and Integral Calculus, vol. 2, Interscience, New York, 1936.
3. N. Dunford and J. Schwartz, Linear Operators, vol. 1, Interscience, New York, 1958.
4. R. A. Frazer, W. J. Duncan, and A. R. Collar, Elementary Matrices, Cambridge University Press, Cambridge, 1938.
5. P. Hilton and S. Wylie, Homology Theory, Cambridge University Press, New York, 1960.
6. M. Hirsch, A proof of the non-retractibility of a cell into its boundary, Proc. Amer. Math. Soc., 14 (1963) 364–365.
7. J. G. Hocking and G. S. Young, Topology, Addison-Wesley, Reading, Mass., 1961.
8. J. Milnor, Analytic proofs of the "Hairy Ball Theorem" and the Brouwer fixed-point theorem, this MONTHLY, 85 (1978) 521–524.
9. T. Muir and W. H. Metzler, A Treatise on the Theory of Determinants, Dover, New York, 1960.
10. H. Scarf, The Computation of Economic Equilibria, Cowles Foundation Monograph 24, Yale University Press, New Haven, 1973.
11. M. Spivak, Calculus on Manifolds, Benjamin, New York, 1965.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

EISENSTEIN RECIPROCITY AND nTH-POWER RESIDUES

JAMES KRAFT AND MICHAEL ROSEN
Department of Mathematics, Brown University, Providence, RI 02912

There is a very pretty theorem about nth power residues, which was discovered first by E. Trost **[5]** and rediscovered later, and independently, by N. C. Ankeny and C. A. Rogers **[1]**. The statement is as follows.

THEOREM 1. *Let a and n be rational integers with n positive. If $x^n \equiv a(p)$ has a solution for all but finitely many primes p then*
(i) *if $8 \nmid n$ we have $a = b^n$, and*
(ii) *if $8|n$ we have $a = b^n$ or $a = 2^{n/2}b^n$.*

It is not difficult to see that $x^8 \equiv 16(p)$ is solvable for all primes p. Thus the second part of the statement is indeed necessary.

The proof of the theorem depends on results about the density of primes that split completely in an algebraic number field. The proof of these results uses analytic methods. In this note we shall prove a special case of Theorem 1 using the Eisenstein reciprocity law, whose proof is purely arithmetic and is, moreover, the easiest of all the higher reciprocity laws. Namely, we shall prove:

THEOREM 2. *Let a and n be rational integers. Suppose that n is positive and square free and that $(a,n) = 1$. If $x^n \equiv a(p)$ has a solution for all but finitely many primes p then $a = b^n$.*

We begin the proof with a brief review of some definitions and the statement of Eisenstein reciprocity.

The rational integers will be denoted by Z. We denote by ζ a primitive lth root of unity, where l is an odd rational prime. Let D be the ring of integers in the field generated over the rationals by ζ. Suppose $P \subset D$ is a prime ideal not containing l. For $\alpha \in D$ we now define the lth-power residue symbol (α/P). If $\alpha \in P$ we set $(\alpha/P) = 0$. If $\alpha \notin P$ we define (α/P) to be the unique lth root of unity such that

$$\alpha^{NP-1/l} \equiv (\alpha/P)(P)$$

where NP is the number of elements in D/P. This symbol is easily seen to be multiplicative in α, and, as the multiplicative group of D/P is cyclic with order divisible by l, we see that $(\alpha/P) = 1$ if and only if $x^l \equiv a(p)$ is solvable, i.e., if and only if a is an lth-power residue modulo p.

Suppose A is an ideal in D that is relatively prime to l and that $A = P_1P_2 \cdots P_t$ is its prime decomposition in D. Then (α/A) is defined to be the product $(\alpha/P_1)(\alpha/P_2)\cdots(\alpha/P_t)$. Finally, if α, $\beta \in D$ and β is prime to l, then (α/β) is defined to be $(\alpha/(\beta))$ where (β) is the principal ideal in D generated by β. The symbol (α/β) is a generalization of the familiar Jacobi symbol from the theory of quadratic residues and enjoys many of the same properties.

Before stating Eisenstein's law of reciprocity we need one more definition. Let $L = (1 - \zeta)$. Observe that L is the unique prime ideal in D lying over the prime l. An algebraic integer is said to be primary if it is congruent to a rational integer modulo L^2.

The second author wishes to thank the Vaughn Foundation for financial support during the period in which this paper was written.

THEOREM 3 (the Eisenstein reciprocity law). *Let a be a rational integer and $\alpha \in D$ a primary algebraic integer. Assume $l \nmid a$. Then $(a/\alpha) = (\alpha/a)$.*

The proof of this theorem can be found in the third volume of E. Landau's book [**3**] and also in the more recent historical essay of A. Weil [**6**]. Unfortunately, we know of no convenient reference in English for this elegant result.

We now begin the proof of Theorem 2. In the case $n = 2$ a proof based solely on the law of quadratic reciprocity can be found in the book of K. Ireland and M. Rosen [**2**, p. 60, Theorem 3], and in the book of W. LeVeque [**5**, p. 122, Theorem 5.10]. A simple reduction now shows it is sufficient to consider the case when $n = l$, an odd prime. For suppose n is square free and $x^n \equiv a(p)$ is solvable for all but finitely many primes p. If $n = l_1 l_2 \cdots l_t$, with the l_i distinct primes, and if we knew the result for prime exponents, it would follow that a is an l_ith power for $i = 1, 2, \ldots, t$. By considering the prime decomposition of a, one then shows easily that a is an nth power. From now on we assume $n = l$, an odd prime, and that $l \nmid a$.

The proof will be established by showing that if a is not an lth power in Z then there are infinitely many rational primes q such that a is not an lth-power residue modulo q.

Assume a is not an lth power in Z and let $(a) = P_1^{e_1} P_2^{e_2} \cdots P_t^{e_t}$ be the prime decomposition of (a) in D. If $l \mid e_i$ for each $i = 1, 2, \ldots, t$, then by taking norms we find $a^{l-1} = b^l$ for some $b \in Z$. It then follows that a is an lth power in Z, contrary to assumption. Thus $l \nmid e_i$ for at least one i. We may assume $l \nmid e_t$.

Let $\{Q_1, Q_2, \ldots, Q_k\}$ be a finite set of prime ideals different from the P_i and from L.

Using the Chinese remainder theorem, we can find an element $\tau \in D$ such that $\tau \equiv 1(Q_i)$ for $i = 1, 2, \ldots, k$, $\tau \equiv 1(L^2)$, $\tau \equiv 1(P_j)$ for $j = 1, 2, \ldots, t-1$, and $\tau \equiv \alpha(P_t)$ where α is chosen so that $(\alpha/P_t) = \zeta$.

Since $\tau \equiv 1(L^2)$, τ is primary, Thus, on the one hand,

$$(a/\tau) = (\tau/a) = \Pi(\tau/P_i)^{e_i} = \zeta^{e_t} \neq 1.$$

On the other hand, let $(\tau) = R_1 R_2 \cdots R_m$ be the prime decomposition of τ in D. Then, by definition,

$$(a/\tau) = \Pi(a/R_j).$$

It follows that for some j we must have $(a/R_j) \neq 1$.

From the congruences that τ satisfies it follows immediately that $R_j \notin \{Q_1, Q_2, \ldots, Q_k\} \cup \{L\} \cup \{P_1, P_2, \ldots, P_t\}$.

We have shown that there are infinitely many prime ideals Q in D such that a is not an lth-power residue modulo Q. Let $qZ = Q \cap Z$. Then a is not an lth-power residue modulo q in Z. Moreover, there are infinitely many such primes q since every rational prime is contained in only finitely many prime ideals in D.

The proof is complete.

References

1. N. C. Ankeny and C. A. Rogers, A conjecture of Chowla, Ann. of Math. (2) 53 (1951) 541–550.
2. K. Ireland and M. Rosen, Elements of Number Theory, Bogden and Quigley, Tarrytown-on-Hudson, N.Y., 1972.
3. E. Landau, Vorlesungen über Zahlentheorie, Chelsea, 1947.
4. W. J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, Reading, Mass., 1978.
5. E. Trost, Zur Theorie der Potenzreste, Nieuw Arch. Wisk., 18 (1934) 58–61.
6. A. Weil, La cyclotomie jadis et naguère, Enseign. Math. (2) 20 (1974) 247–263.

AN EASY PROOF OF THE TRIPLE-PRODUCT IDENTITY

JOHN A. EWELL

Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115

1. It is the purpose of this note to derive the celebrated Gauss-Jacobi triple-product identity

as an easy consequence of Euler's pentagonal-number identity and a theorem of Gauss. The last two results are stated as follows.

THEOREM 1. *For each complex number* x *with* $|x| < 1$,

$$\prod_{n=1}^{\infty} (1 - x^n) = \sum_{n=-\infty}^{\infty} (-1)^n x^{n(3n+1)/2}. \tag{1}$$

THEOREM 2. *For each pair of complex numbers* a, x, *with* $a \neq 0$ *and* $|x| < 1$,

$$\prod_{n=1}^{\infty} (1 + ax^{2n-1})(1 + a^{-1}x^{2n-1}) = P_0(x) \sum_{n=-\infty}^{\infty} a^n x^{n^2}, \tag{2}$$

$P_0(x)$ *an undetermined function of* x.

For an accessible proof of Theorem 1, see the elegant combinatorial argument of F. Franklin, as given in [**1**, p. 286] or [**2**, p. 21]. Since Gauss's argument is perhaps less well known, we give a modernized reproduction of it in the following section. For each pair of positive real numbers A, X, with $X < 1$, the infinite product $F(a, x)$ on the left of (2) converges absolutely and uniformly for all pairs a, x such that $A^{-1} \leqslant |a| \leqslant A$ and $|x| \leqslant X$. For a fixed choice of $x, |x| < 1$, $F(a, x)$ thus defines a unique function of a, which is analytic at all points of the finite complex plane except $a = 0$ where it has an essential singularity. Accordingly, $F(a, x)$ has a unique Laurent series expansion about 0. Of course, the coefficients of this series are then functionally dependent on the chosen x.

2. Proof of Theorem 2. On the strength of the final statement of section 1, we express the infinite product $F(a, x)$ as follows:

$$F(a, x) = P_0(x) + \sum_{n=1}^{\infty} \{P_n(x)a^n + P_{-n}(x)a^{-n}\}.$$

Since $F(a, x) = F(a^{-1}, x)$, it follows that, for each positive integer n, $P_n(x) = P_{-n}(x)$, whence

$$F(a, x) = P_0(x) + \sum_{n=1}^{\infty} P_n(x)\{a^n + a^{-n}\}. \tag{3}$$

We now easily establish the relation $F(ax^2, x) = (ax)^{-1}F(a, x)$, then use the series representations obtained from (3) to expand both sides of this identity in powers of a and equate coefficients of like powers to derive the following recurrence: $P_n(x) = P_{n-1}(x)x^{2n-1}$. Iteration of the foregoing yields: $P_n(x) = P_0(x)x^{n^2}$ for each nonnegative integer n. Substituting these values into (3), we thus prove Theorem 2.

3. According to MacMahon [**2**, p. 78], Gauss proved Theorem 2 and was, up to evaluation of $P_0(x)$, at the threshold of the triple-product identity. However, it apparently escaped his attention that Euler's identity (1) provides an easy way of evaluating $P_0(x)$. For, in (2) we replace a and x by $-x$ and x^3, respectively, to get

$$\begin{aligned} \prod_{n=1}^{\infty} (1 - x^{2(3n-1)})(1 - x^{2(3n-2)}) &= P_0(x^3) \sum_{n=-\infty}^{\infty} (-1)^n x^{n(3n+1)} \\ &= P_0(x^3) \prod_{n=1}^{\infty} (1 - x^{2n}). \end{aligned}$$

[In the last step we have used Euler's identity.] Cancellation of the factors corresponding to the residues 1 and 2 (mod 3) yields:

$$P_0(x^3) \prod_{n=1}^{\infty} (1 - x^{6n}) = 1,$$

whence

$$P_0(x) = \prod_{n=1}^{\infty} (1 - x^{2n})^{-1}.$$

The foregoing argument tacitly assumes that $x \neq 0$. However, for $x = 0$ we appeal directly to (2) taking $a = 1$, say, and conclude that $P_0(0) = 1$; and, of course, $\prod_{n=1}^{\infty}(1 - 0^{2n})^{-1} = 1$, as well. Thus, we have proved the following:

THEOREM 3. *For each pair of complex numbers a, x, with $a \neq 0$ and $|x| < 1$,*

$$\prod_{n=1}^{\infty} (1 - x^{2n})(1 + ax^{2n-1})(1 + a^{-1}x^{2n-1}) = \sum_{n=-\infty}^{\infty} a^n x^{n^2}.$$

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960.

2. P. A. MacMahon, Combinatory Analysis, vol. 2, Chelsea, New York, 1960.

OCCURRENCES OF CONSECUTIVE ODD POWERFUL NUMBERS

W. A. SENTANCE

Department of Computer Science, The City University, Northampton Square, London EC1V OHB, England

S. W. Golomb [1] defined a *powerful number* to be a positive integer r such that p^2 divides r whenever the prime p divides r and posed the question whether the pair (25, 27) is the only occurrence of consecutive odd powerful numbers. In this note we show that there exists an infinite number of such pairs, each pair being associated with a certain Pell equation. This association arises from the following theorem and its converse.

THEOREM. *If the equation $x^2 - my^2 = 1$ has a solution in which* (1) *x is even, and* (2) *m divides y (implying that my^2 is powerful), then $(x - 1, x + 1)$ are consecutive odd powerful numbers.*

Proof. We write

$$(x + 1)(x - 1) = x^2 - 1 = my^2.$$

Since x is even, $x + 1, x - 1$ are consecutive odd numbers and are therefore relatively prime. Their product my^2 is powerful and so each of $x - 1, x + 1$ is powerful in its own right.

The case $x = 26$, $y = 15$, $m = 3$ satisfies the conditions and gives rise to the pair (25, 27) quoted earlier.

CONVERSE. *If $x - 1, x + 1$ are consecutive odd powerful numbers, then x, y is the solution of some equation $x^2 - my^2 = 1$, where m divides y; it is obvious that x is even.*

Proof. Since

$$x^2 - 1 = (x - 1)(x + 1)$$

is powerful, we can write

$$x^2 - 1 = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

where the p_i's are distinct primes and $n_i \geqslant 2$. Without loss of generality, let

$$n_i \equiv 1 \pmod 2 \qquad 1 \leqslant i \leqslant l \leqslant k,$$
$$n_i \equiv 0 \pmod 2 \qquad l + 1 \leqslant i \leqslant k.$$

Then

$$\begin{aligned} x^2 - 1 &= p_1 p_2 \cdots p_l (p_1^{n_1 - 1} p_2^{n_2 - 1} \cdots p_l^{n_l - 1}) p_{l+1}^{n_{l+1}} \cdots p_k^{n_k} \\ &= my^2 \quad \text{where } m (= p_1 p_2 \cdots p_l) \text{ divides } y. \end{aligned}$$

The solution of the Pell equation $x^2 - my^2 = 1$ is closely linked with the continued-fraction representation of $\sqrt{m}$.

If $\sqrt{m}$ is irrational it has a continued fraction representation

$$\sqrt{m} = [a_0, \overline{a_1, a_2, \ldots, a_2, a_1, 2a_0}].$$

The a_i's are all integers, and a simple algorithm for their determination is given in [**2**, § 10.6]. The terms under the bar form a recurrent sequence and their number, n, is the period of the continued fraction.

If we construct two sequences p_i, q_i defined by

$$\begin{array}{ll} p_0 = a_0 & q_0 = 1 \\ p_1 = a_1 p_0 + 1 & q_1 = a_1 \\ p_2 = a_2 p_1 + p_0 & q_2 = a_2 q_1 + q_0 \\ \vdots & \vdots \\ p_i = a_i p_{i-1} + p_{i-2} & q_i = a_i q_{i-1} + q_{i-2}, \end{array}$$

then the values $x = p_{kn-1}, y = q_{kn-1}, k = 1,2,\ldots,$ satisfy the Pell equation

$$x^2 - my^2 = (-1)^{kn}.$$

We now show that in the case $n = 2$ certain solutions of $x^2 - my^2 = 1$ satisfy the given conditions of our theorem.

The general solution for $n = 2$ is given by $x = p_{2l-1}$, $y = q_{2l-1}$, $l = 1,2,\ldots,$ and the continued fraction expansion of $\sqrt{m}$ is $[a, \overline{b, 2a}]$, where $m = a^2 + 2a/b$. The first few values of the p_i sequence are

$$\begin{aligned} p_0 &= a \\ p_1 &= bp_0 + 1 \\ p_2 &= 2ap_1 + p_0 \\ p_3 &= bp_2 + p_1, \end{aligned}$$

which for odd i satisfy the difference equation

$$p_{2l+3} = 2(ab+1)p_{2l+1} - p_{2l-1}.$$

Using congruences modulo 2 we distinguish two cases.

(1) $l = 2s, p_{4s+3} \equiv p_{4s-1} \equiv \cdots \equiv p_3 = 2ab(ab+2) + 1.$

Hence p_{4s+3} is always odd.

(2) $l = 2s - 1, p_{4s+1} \equiv p_{4s-3} \equiv \cdots \equiv p_1 = ab + 1.$

Hence by choosing a and b both odd, p_{4s+1} is even and we have solutions $x = p_{4s+1}, y = q_{4s+1}$, $s = 1,2,\ldots,$ for which x is even, thus satisfying our first condition. We find the members of this family of solutions for which m divides y, that is, q_{4s+1}, by considering the difference equation satisfied by the q_i sequence for odd i. This is

$$q_{2l+3} - 2(c+1)q_{2l+1} + q_{2l-1} = 0$$

where $c = ab + 1$. This has a general solution

$$q_{2l+1} = A(c + b\sqrt{m})^l + B(c - b\sqrt{m})^l.$$

We determine A and B from the initial conditions

$$\begin{array}{ll} l = 0, & q_1 = b, \\ l = 1, & q_3 = 2bc, \end{array}$$

giving the particular solution

$$q_{2l+1} = \frac{1}{2\sqrt{m}}\left[(c + b\sqrt{m})^{l+1} - (c - b\sqrt{m})^{l+1}\right].$$

If we substitute $l = 2s$, then

$$q_{4s+1} = \frac{1}{2\sqrt{m}}\left[(c + b\sqrt{m})^{2s+1} - (c - b\sqrt{m})^{2s+1}\right].$$

Expanding $(c \pm b\sqrt{m})^{2s+1}$, using the binomial theorem, we obtain

$$q_{4s+1} = \frac{1}{2\sqrt{m}} \sum_{r=0}^{2s+1} c^{2s+1-r}\binom{2s+1}{r}\left[(b\sqrt{m})^r - (-b\sqrt{m})^r\right],$$

$$= \frac{1}{2\sqrt{m}} \sum_{r=0}^{2s+1} c^{2s+1-r}\binom{2s+1}{r}2(b\sqrt{m})^r,$$

where the summation is over odd values of r only,

$$= \frac{1}{2\sqrt{m}}\left[c^{2s}\binom{2s+1}{1}2b\sqrt{m} + \text{terms in } m\sqrt{m}, m^2\sqrt{m}, \text{etc.}\right].$$

Therefore $q_{4s+1} = bc^{2s}(2s + 1)$ + terms in m, m^2, etc. Since

$$c^2 = (ab + 1)^2 = a^2b^2 + 2ab + 1 = b^2m + 1,$$

$$q_{4s+1} \equiv b(b^2m + 1)^s(2s + 1) \pmod{m},$$

$$\equiv b(2s + 1) \pmod{m}.$$

If $t = (b, m)$ is the highest common factor of b and m, then $b = tg$ and $m = th$ where g and h are relatively prime. For m to divide q_{4s+1}, the expression $g(2s + 1)/h$ must be integral. Therefore $2s + 1$ needs to be an integral multiple of h. This is possible, because, for a being odd, $m(= a^2 + 2a/b)$ is odd and hence h is odd. Thus for values of s that satisfy $2s + 1 = (2j + 1)h, j = 0, 1, 2, \ldots,$ m divides q_{4s+1}. Corresponding to the value of j the subscript $4s + 1$ will take the value $4jh + 2h - 1$. This establishes the main result of this note:

THEOREM. *If $\sqrt{m} = [a, \overline{b, 2a}]$, where a and b are both odd and $(b, m) = t$, then the solutions of $x^2 - my^2 = 1$ given by $x = p_l, y = q_l$ for $l = (4j + 2)m/t - 1, j = 0, 1, 2, \ldots,$ satisfy the conditions of our earlier theorem. Thus $p_l \pm 1$ is a pair of consecutive odd powerful numbers.*

Applying our result, for example, to the equation $x^2 - 3y^2 = 1$, we have for $m = 3$ that $\sqrt{3} = [1, \overline{1, 2}]$ giving $a = b = t = 1$. The general solution is therefore $p_l \pm 1$ for $l = 12j + 5$. The first three pairs are

$j = 0$, $p_5 \pm 1 = (25, 27)$—the pair quoted by Golomb,

$j = 1$, $p_{17} \pm 1 = (70225, 70227)$,

$j = 2$, $p_{29} \pm 1 = (189750625, 189750627)$.

Consecutive odd powerful numbers are relatively prime. Thus we have shown also that 2 has an infinite number of representations as a difference of two relatively prime powerful numbers. Such a representation is termed a *proper representation*. This complements the work of Golomb who showed that 1 and 4 have this property and of Makowski [3] who showed it to be true for every prime $p \equiv 1 \pmod 8$.

References

1. S. W. Golomb, Powerful numbers, this MONTHLY, 77 (1970) 848–852.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, 1960.
3. A. Makowski, On a problem of Golomb on powerful numbers, this MONTHLY, 79 (1972) 761.

NOTE ON AN INTEGRAL INEQUALITY OF KY FAN AND G. G. LORENTZ

JOSIP E. PEČARIĆ
Faculty of Civil Engineering, Bulevar Revolucije 73, 11000 Beograd, Yugoslavia

Ky Fan and G. G. Lorentz [1] (see also [2, p. 304]) have proved the following result:

THEOREM A. *Let H be a real function depending on t and $u_1,\dots,u_n$, defined for $0 \leqslant t \leqslant 1$, $-\infty < u_i < +\infty$, for $i = 1,\dots,n$, and having continuous second derivatives. For*

$$\int_0^1 H(t;f_1(t),\dots,f_n(t))\,dt \leqslant \int_0^1 H(t;g_1(t),\dots,g_n(t))\,dt \tag{1}$$

to hold for each system of decreasing bounded functions f_i, g_i $(i = 1,\dots,n)$ satisfying

$$\int_0^x f_i(t)\,dt \leqslant \int_0^x g_i(t)\,dt \qquad (0 \leqslant x \leqslant 1, 1 \leqslant i \leqslant n)$$

and

$$\int_0^1 f_i(t)\,dt = \int_0^1 g_i(t)\,dt \qquad (1 \leqslant i \leqslant n),$$

it is necessary and sufficient that

$$\frac{\partial^2 H}{\partial u_i \partial u_j} \geqslant 0 \quad (i,j = 1,\dots,n) \quad \text{and} \quad \frac{\partial^2 H}{\partial t \partial u_i} \leqslant 0 \quad (i = 1,\dots,n). \tag{2}$$

We shall give a short proof of a more general result:

THEOREM 1. *Let H be a real function depending on t and $u_1,\dots,u_n$, defined for $0 \leqslant t \leqslant 1$, $-\infty < u_i < +\infty$, for $i = 1,\dots,n$, and having continuous second derivatives. Let f_i, g_i $(1 \leqslant i \leqslant n)$ be nonincreasing bounded functions, and let p be an integrable real function. Let a function*

$$F(\lambda) = \int_0^1 p(t)H(t;u_1(t;\lambda),\dots,u_n(t;\lambda))\,dt$$

where $u_i(t;\lambda) = \lambda g_i(t) + (1-\lambda)f_i(t)$ $(1 \leqslant i \leqslant n)$, be defined for $0 \leqslant \lambda \leqslant 1$.

(a) *If* (2) *holds, and if*

$$\int_0^x p(t)f_i(t)\,dt \leqslant \int_0^x p(t)g_i(t)\,dt \qquad (0 \leqslant x \leqslant 1, 1 \leqslant i \leqslant n),$$

$$\int_0^1 p(t)f_i(t)\,dt = \int_0^1 p(t)g_i(t)\,dt \qquad (1 \leqslant i \leqslant n), \tag{3}$$

then $F(\lambda)$ is a nondecreasing function on $[0,1]$.

(b) *If* (2) *and $\partial H/\partial u_i \geqslant 0$ $(1 \leqslant i \leqslant n)$ hold, and if*

$$\int_0^x p(t)f_i(t)\,dt \leqslant \int_0^x p(t)g_i(t)\,dt \qquad (0 \leqslant x \leqslant 1, 1 \leqslant i \leqslant n), \tag{4}$$

then $F(\lambda)$ is a nondecreasing function, too.

Proof. The functions $u_i(t;\lambda)$ are nonincreasing on $0 \leqslant t \leqslant 1$, so that

$$\begin{aligned} F'(\lambda) &= \sum_{i=1}^n \int_0^1 p(t)(g_i(t) - f_i(t)) \frac{\partial H(t;u_1(t;\lambda),\dots,u_n(t;\lambda))}{\partial u_i}\,dt \\ &= \sum_{i=1}^n \frac{\partial H(1;u_1(1;\lambda),\dots,u_n(1;\lambda))}{\partial u_i} \int_0^1 p(t)(g_i(t) - f_i(t))\,dt \\ &\quad - \sum_{i=1}^n \int_0^1 \left(\int_0^x p(t)(g_i(t) - f_i(t))\,dt\right) d\left(\frac{\partial H(x;u_1(x;\lambda),\dots,u_n(x;\lambda))}{\partial u_i}\right) \\ &\geqslant 0, \end{aligned}$$

because

$$\frac{d}{dt}\left(\frac{\partial H(t;u_1(t;\lambda),\dots,u_n(t;\lambda))}{\partial u_i}\right)=\frac{\partial^2 H(t;u_1(t;\lambda),\dots,u_n(t;\lambda))}{\partial t\partial u_i}$$

$$+\sum_{j=1}^{n}\frac{\partial^2 H(t;u_1(t;\lambda),\dots,u_n(t;\lambda))}{\partial u_i\partial u_j}\frac{du_j(t;\lambda)}{dt}\leqslant 0.$$

Thus Theorem 1 is proved.

COROLLARY 2. *If the conditions of Theorem* 1 *are satisfied and if* $0\leqslant\alpha<\beta\leqslant 1$, *then*

$$\int_0^1 p(t)H(t;f_1(t),\dots,f_n(t))\,dt\leqslant\int_0^1 p(t)H(t;u_1(t;\alpha),\dots,u_n(t;\alpha))\,dt$$
$$\leqslant\int_0^1 p(t)H(t;u_1(t;\beta),\dots,u_n(t;\beta))\,dt\leqslant\int_0^1 p(t)H(t;g_1(t),\dots,g_n(t))\,dt. \tag{5}$$

Inequalities (5) are refinements of the following inequality

$$\int_0^1 p(t)H(t;f_1(t),\dots,f_n(t))\,dt\leqslant\int_0^1 p(t)H(t;g_1(t),\dots,g_n(t))\,dt, \tag{6}$$

which is the generalization of (1).

However, we can obtain (5) from (6). Indeed, denote (3) and (4) by $f_i\prec g_i$ and $f_i\prec\prec g_i$; then

$$f_i\prec g_i\Leftrightarrow u_i(t;\alpha)\prec u_i(t;\beta)\quad\text{and}\quad f_i\prec\prec g_i\Leftrightarrow u_i(t;\alpha)\prec\prec u_i(t;\beta)$$

where $0\leqslant\alpha<\beta\leqslant 1$. Using the substitutions $f_i(t)=u_i(t;A)$ and $g_i(t)=u_i(t;B)$ in the cases: (i) $A=0$, $B=\alpha$; (ii) $A=\alpha$, $B=\beta$; (iii) $A=\beta$, $B=1$; from (6) we obtain (5).

The author is grateful to Professors Ky Fan and P. M. Vasić for useful suggestions.

References

1. Ky Fan and G. G. Lorentz, An integral inequality, this MONTHLY 61 (1954) 616–631.

2. D. S. Mitrinović (in cooperation with P. M. Vasić), Analytic Inequalities, Springer, Berlin-Heidelberg-New York, 1970.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

SHIFTED-PRIME FACTORIZATIONS

KENNETH H. ROSEN

Department of Mathematics, University of Maine, Orono, ME 04469

A **completely additive arithmetic function** f is a mapping of the positive integers to the complex numbers such that $f(ab)=f(a)+f(b)$ for all positive integers a and b. Kátai [5] conjectured and P. D. T. A. Elliott [1] proved that, if f is a completely additive arithmetic function such that $f(p+1)=0$ for all primes p, then f must be identically zero. An immediate consequence of this

fact, established by Wolke [7] is the surprising result that every positive integer may be expressed in the form

$$n = \prod_{i=1}^{u} (p_i + 1)^{r_i}$$

where the p_i are primes and the exponents r_i are rational numbers. Here such representations will be called **shifted-prime factorizations.** Such expansions for an integer are not unique; one has for instance that $2 = (3 + 1)^{1/2} = (31 + 1)^{1/5}$.

There are many questions that arise concerning shifted-prime factorizations. For example, if $R(n)$ is the minimum value, over all shifted-prime factorizations of n, of the largest prime occurring in the factorization, one may ask:

Question 1. *Find an upper bound for $R(n)$.*

No bounds for $R(n)$ are known, but using an unproved conjecture of Kanold [3], [4], and Schinzel and Sierpiński [6] one can provide an answer to Question 1.

CONJECTURE 1. *Let a and b be relatively prime positive integers. Then there is a prime p with $p < a^2$ in the arithmetic progression $an + b$, $n = 0, 1, 2, \ldots$.*

For the remainder of this note p will always represent a prime. Let $Q(p) = pm(p) - 1$ be the least prime in the arithmetic progression $pn - 1$. A result that follows from Conjecture 1 is

$$R(n) \leqslant \max_{p<n} Q(p) < \max_{p<n} p^2$$

for all positive integers n. To show that this assertion is a consequence of Conjecture 1 use induction. Note that $Q(2) = 3$ and $2 = (3 + 1)^{1/2}$, so that $R(2) = 3$. Now assume that $R(k) \leqslant \max_{p<k} Q(p) < \max_{p<k} p^2$ for all positive integers $k < n$. If n is composite, then $n = st$ with $1 < s \leqslant t < n$. By the induction hypothesis

$$R(n) \leqslant \max\Big(\max_{p<s} Q(p), \max_{p<t} Q(p)\Big) \leqslant \max_{p<n} Q(p) < \max_{p<n} p^2.$$

When n is prime, write $n = (Q(n) + 1)m(n)^{-1}$. From Conjecture 1 it follows that $Q(n) < n^2$ and $m(n) < n$ (note that $m(n) \neq n$ since $n^2 - 1$ is composite). By the inductive hypothesis $R(m(n)) \leqslant \max_{p<m(n)} Q(p) \leqslant \max_{p<n} Q(p)$. Hence one concludes that

$$R(n) \leqslant \max(Q(n), R(m(n))) \leqslant \max_{p<n} Q(p) < \max_{p<n} p^2,$$

completing the induction argument.

A second related problem is:

Question 2. *Give an upper bound for the minimum number of primes required to represent all positive integers not exceeding n by shifted-prime factorizations involving only these primes.*

By the argument above one sees that Conjecture 1 implies that for each successive prime p one need only add the prime $Q(p)$ to the set of primes used in shifted-prime factorizations. Hence if Conjecture 1 is valid one would need less than or equal to $\pi(n)$ primes, where $\pi(n)$ is the number of primes not exceeding n, to represent all positive integers less than or equal to n. Therefore Conjecture 1 has as a consequence:

CONJECTURE 2. *Every positive integer n may be written as*

$$n = \prod_{i=1}^{v} (Q(p_i) + 1)^{r_i}$$

where each p_i is prime, $Q(p_i)$ is the least prime of the form $np_i - 1$, and the exponents r_i are rational numbers.

An equivalent form of Conjecture 2 is

CONJECTURE 2′. *If f is a completely additive arithmetic function such that $f(Q(p)+1)=0$ for every prime p, then f is identically zero.*

Here are expansions of small primes of the type required by Conjecture 2.

$$
\begin{array}{ll}
2=(3+1)^{1/2} & 13=(103+1)(3+1)^{-3/2} \\
3=2+1 & 17=(67+1)(3+1)^{-1} \\
5=(19+1)(3+1)^{-1} & 19=(37+1)(3+1)^{-1/2} \\
7=(13+1)(3+1)^{-1/2} & 23=(137+1)(3+1)^{-1/2}(2+1)^{-1} \\
11=(43+1)(3+1)^{-1} & 29=(173+1)(3+1)^{-1/2}(2+1)^{-1}
\end{array}
$$

If established, the following conjecture of P. D. T. A. Elliott [2] would provide the best possible answer to Question 2.

CONJECTURE 3. *Every nonzero rational number can be expressed in the form $(p+1)/(q+1)$ where p and q are primes (a negative prime being allowed for negative rationals).*

The author was supported in part by NSF Grant MCS 77-02171.

References

1. P. D. T. A. Elliott, A conjecture of Kátai, Acta Arith., 26 (1974) 11–20; **MR** 50 #7077.
2. ———, personal communication.
3. H.-J. Kanold, Elementare Betrachtungen zur Primzahltheorie, Arch. Math., 14 (1963) 147–151; **MR** 27 #89.
4. ———, Über Primzahlen in arithmetischen Folgen, Math. Ann., 156 (1964) 393–395; **MR** 30 #70.
5. I. Kátai, On sets characterizing number-theoretical functions, Acta Arith., 13 (1968) 315–320; **MR** 36 #3742.
6. A Schinzel and W. Sierpiński, Sur certaines hypothèses concernant les nombres premiers, Acta Arith., 4 (1958) 185–208; **MR** 21 #4936.
7. D. Wolke, Bemerkungen über Eindeutigkeitsmengen additiver Funktionen, Elem. Math., 33 (1978) 14–16; **MR** 57 #3058.

FANOUT-FREE FUNCTIONS

EDWARD A. BENDER
Department of Mathematics, University of California–San Diego, La Jolla, CA 92093

E. RODNEY CANFIELD
Department of Computer Science, University of Georgia, Athens, GA 30602

In computer science parlance, a **gate** is a function $g\colon S^n \to S$ for some finite set S and positive integer n. We say that g has n **inputs** and one **output**, the function value. Gates can be connected together by using outputs of some gates as inputs of others. For example, if the inputs x, y, z, w and the gates f, g are connected as shown in Fig. 1, the resulting **network** computes the function $g(x, f(y,z), w)$. If the original inputs appear only once in this function decomposition, the network is called **fanout-free** and can be represented by a rooted tree with the inputs as leaves, the gates as nodes, and the output as root.

The problem is this: *Find a canonical form theorem for functions realizable as fanout-free networks using a specified collection of gates.*

By "a canonical form theorem" we mean a set of restrictions such that any function realizable in one way or in many different ways as a fanout-free network using the specified gates has precisely one realization as a network satisfying the imposed restrictions. We will

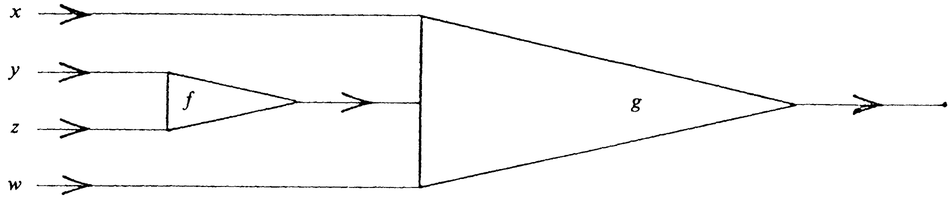


FIG. 1

illustrate this with a specific example later, and the reader is probably familiar with the notion of "canonical form" from matrix theory.

Since functions do not generally admit unique decompositions as fanout-free networks, one cannot enumerate the functions that can be synthesized from a specified set of gates simply by enumerating the fanout-free networks whose nodes are selected from the given gates. However, [**1**] shows how this difficult enumeration problem is resolved using a canonical form theorem. A canonical form theorem may also be useful in developing a synthesis algorithm, that is, one that finds a representation of a given function as a fanout-free network [**4**]. See [**3**] for more details on why fanout-free networks are of interest.

We now describe the canonical form when $S = \mathrm{GF}(2)$, the two-element finite field. We use the symbol $+$ to denote the usual operation of addition in this field, for which $1 + 1 = 0$. The functions $f_1(x,y) = xy$ and $f_2(x,y) = x + y$, called AND and XOR, are a source of difficulty because $f_i(x,f_i(y,z)) = f_i(f_i(x,y),z)$. In more standard algebraic terms, the operation $x*y = f_i(x,y)$ is associative. The complementation operation $c(x) = x + 1$ is a bit of a problem, too, because $c(c(x)) = x$ and $c(f_2(x,y)) = f_2(c(x),y)$. These are essentially the only problems that prevent a function from having a unique decomposition, as illustrated by the following result.

THEOREM. *Let G be a set of gates such that*:

(1) *if $g \in G$, then $g(\mathbf{0}) = 0$ and, for every $\mathbf{d} \in S^n$, the gate $h(\mathbf{x}) = g(\mathbf{d} + \mathbf{x}) + g(\mathbf{d})$ is in G (we call g and h equivalent*);

(2) *if $a_n(\mathbf{x}) = x_1x_2 \cdots x_n \in G$ for some $n \geqslant 2$, then $a_m \in G$ for all $m \geqslant 2$*;

(3) *if $s_n(\mathbf{x}) = x_1 + x_2 + \cdots + x_n \in G$ for some $n \geqslant 2$, then $s_m \in G$ for all $m \geqslant 2$*;

(4) *with the exception of a_n, s_n, and $h(\mathbf{x}) = a_n(\mathbf{x} + \mathbf{d}) + a_n(\mathbf{d})$, no $g \in G$ with n inputs is the same function as a fanout-free network using only gates with less than n inputs.*

Then every function that can be synthesized by a fanout-free network using G and complementation has a unique synthesis using G and complementation and satisfying the following rules:

(a) *if complementation is used at all, it is the output gate*;

(b) *the output of an s_m is never the input of an s_k for any $m, k \geqslant 2$*;

(c) *if $a_m(\mathbf{x} + \mathbf{d}) + a_m(\mathbf{d})$ is the ith input to $g(\mathbf{x}) = a_k(\mathbf{x} + \mathbf{e}) + a_k(\mathbf{e})$, then $a_m(\mathbf{d}) \neq e_i$.*

The proof of this theorem is implicit in [**4**, Chap. 4].

In attempting to generalize these ideas to logic with more than two values, it might be helpful to limit attention to symmetric gates; i.e., gates g for which $g(\pi(\mathbf{x})) = g(\mathbf{x})$ for every permutation of the components of $\mathbf{x}$. Unfortunately, G necessarily contains gates which are not symmetric. If the gates in G are all equivalent to symmetric gates, one can easily obtain a synthesis with only symmetric gates as follows. Reduce G by eliminating gates equivalent to gates that are retained, add the complementation gate, and call the new set H. The set H contains only symmetric gates. Replace rule (a) by (a′) and add rule (d):

(a′) *the output of complementation is never the input of complementation or s_m*;

(d) *if $h \in H$ satisfies $h(\mathbf{x}) = h(\mathbf{x} + \mathbf{1}) + h(\mathbf{1})$ and $h \neq a_m, s_m$, then the first input of h is zero when all network inputs are zero.*

This canonical form can be deduced from the previous one fairly easily.

The work of the first author was supported by NSF Grant MCS77-02113, and that of the second author by NSF Grant MCS78-01794.

References

1. E. A. Bender, The number of fanout-free functions with various gates, J. Assoc. Comput. Mach., 27 (1980) 181–190.

2. E. A. Bender and J. T. Butler, Enumeration of functions realized by fanout-free networks of general multi-valued gates, Proc. Intl. Symp. Multi-Valued Logic (1979), 94–103.

3. J. T. Butler, On the number of functions realized by cascades and disjunctive networks, IEEE Trans. Comput., C-24 (1975) 681–690.; MR54 4831.

4. H. A. Curtis, A New Approach to the Design of Switching Circuits, Van Nostrand, 1962, Chapter 4.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis MO 63121.

AN ALTERNATIVE TO EUCLID'S ALGORITHM

DANIEL A. MARCUS
Department of Mathematics, California State Polytechnic University, Pomona, CA 91768

Euclid's algorithm for constructing the greatest common divisor of two integers plays at least two important roles in the standard exposition of elementary number theory. Besides serving a practical purpose for solving numerical problems, it also provides the theoretical basis for the Unique Factorization Theorem. Yet the algorithm has some undesirable features. In its general form, it involves a fair amount of obscure subscripted notation. As a practical procedure, it often requires cumbersome algebraic manipulations that are highly prone to errors and are not well suited to efficient implementation on a calculator. This note will present an alternative approach that accomplishes the same results as Euclid's algorithm but which is free from the drawbacks described above.

Euclid's Algorithm. Let m and n be two positive integers. Euclid's algorithm constructs the greatest common divisor (GCD) d of m and n along with integers a and b such that $am + bn = d$. The well-known procedure involves repeated division, resulting in a sequence of equations

$$\begin{aligned} m &= q_1 n + r_1 \quad (0 \leqslant r_1 < n) \\ n &= q_2 r_1 + r_2 \quad (0 \leqslant r_2 < r_1) \\ r_1 &= q_3 r_2 + r_3 \quad (0 \leqslant r_3 < r_2) \\ &\vdots \end{aligned}$$

in which the r_i and q_i are integers. The process terminates when some remainder r_k is equal to 0, which must happen eventually. Then the GCD is r_{k-1}, and the coefficients a and b are found as follows: Solve the first equation for r_1 and substitute into the second and third equations to eliminate r_1; then solve the new second equation for r_2 and substitute into the third and fourth equations to eliminate r_2; etc. Eventually an equation of the form $am + bn = r_{k-1}$ is obtained, where a and b are integers.

The Alternative. The existence of a and b when m and n are relatively prime integers plays a crucial role in the proof of the Unique Factorization Theorem; see, for example, [**2**, p. 26]. It is

known, however, that this existence can be established by a simple nonconstructive argument that does not depend on Euclid's algorithm. (See [**2**, p. 26, Exercise 12], [**1**, p. 20], or [**3**, p. 7].) The argument runs as follows: Given relatively prime integers m and n, let S denote the set of all integers of the form $am + bn$, for all integers a and b. It is clear that S is closed under both addition and subtraction. From these properties it follows that, if s is the least positive member of S, then S contains all multiples of s and no other integers. Since m is clearly a member of S, s must be a divisor of m. Similarly, s is a divisor of n. Since m and n are assumed to be relatively prime, we conclude that $s = 1$. Thus $am + bn = 1$ for some integers a and b.

As for the practical function of Euclid's algorithm, we note first that when all that is required is the GCD of two positive integers m and n then there is no need to compute the quotients q_i, and consequently the language and techniques of modular arithmetic can be utilized: Reduce the larger of m and n modulo the smaller and let the reduced value replace the larger number. Repeat this procedure until 0 is obtained; the last nonzero value is the GCD.

What is less obvious is that the coefficients a and b in the equation $am + bn = d$ can also be obtained easily without computing the q_i. The first step is to perform the sequence of reductions described above. With m denoting the larger of the two numbers (which we will assume from now on), the results are arranged in the form

$$\begin{matrix} m & m_1 & m_1 & m_2 & m_2 & \cdots \\ n & n & n_1 & n_1 & n_2 & \cdots \end{matrix}$$

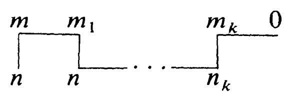
where m reduces to m_1 modulo n, n reduces to n_1 modulo m_1, etc. The process terminates in one of two ways:

$$\begin{matrix} \cdots & m_k & 0 \\ \cdots & n_k & n_k, \end{matrix} \qquad \text{(case 1)}$$

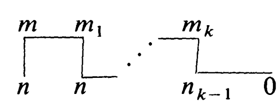
in which case $d = n_k$, or

$$\begin{matrix} \cdots & m_k & m_k \\ \cdots & n_{k-1} & 0, \end{matrix} \qquad \text{(case 2)}$$

in which case $d = m_k$. In either case, construct a path from n to 0 consisting of alternating vertical and horizontal segments as shown:



(case 1)

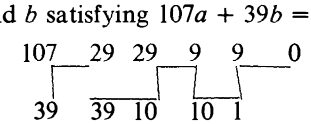


(case 2).

Now follow the path backwards from 0 to n, taking 0 as the initial value and performing the operations indicated by

$$\underset{\times}{\leftarrow} \qquad\qquad \downarrow +d)\ \div \qquad\qquad \uparrow -d)\ \div$$

on the numbers encountered along the way. That is, when moving to the left along the path, multiply by the next number on the path; when moving downward, add d and then divide by the next number; when moving upward, subtract d and divide by the next number. The result at the end of the path is the coefficient b, from which a can then be obtained easily. As an illustration of this procedure, we obtain a and b satisfying $107a + 39b = 1$.



The sequence of operations is

$$0 \times 9) + 1) \div 1) \times 10) - 1) \div 9) \times 29) + 1) \div 10) \times 39) - 1) \div 29) \times 107) + 1) \div 39 = 11,$$

where the first member of each pair of parentheses is understood to occur at the beginning of the calculation. Thus $b = 11$, and consequently $a = -4$.

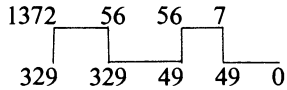
The sequence of operations above is carried out quickly and easily on a calculator, without any need to store data along the way. There is also no need to write down the operations as we did above, since they can be read directly off the path.

No rounding errors occur because all numbers computed turn out to be integers. This fact, which also provides a check on the calculations, will be established along with the validity of the algorithm.

The algorithm can be expressed in a particularly compact form in reverse Polish notation, utilizing the symbol "\", where "$a \backslash b$" means "b/a":

$$\text{Case 1:}\quad n\,m \ldots 0 * d + \backslash * d - \backslash \ldots * d + \backslash$$
$$\text{Case 2:}\quad n\,m \ldots 0 * d - \backslash * d + \backslash \ldots * d + \backslash$$

As another illustration, we compute a and b satisfying $1372a + 329b = 7$.

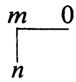


$$0 \times 49) - 7) \div 7) \times 56) + 7) \div 49) \times 329) - 7) \div 56) \times 1372) + 7) \div 329 = -25.$$

Thus $b = -25$, and consequently $a = 6$.

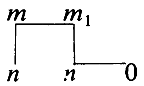
The proof that this algorithm gives the correct value of b and that all numbers computed are integers can be accomplished by induction on the number of steps, which we define to mean the number of horizontal segments in the path from 0 to n.

Suppose the algorithm requires only one step.



Then the GCD is n and the calculation correctly gives $0 \times m) + n) \div n = 1$ as the value of b.

Next, suppose the algorithm ends in two steps.



Then the GCD is m_1 and the calculation gives

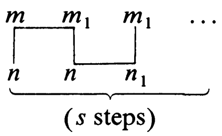
$$0 \times n) - m_1) \div m_1) \times m) + m_1) \div n = (-m + m_1) \div n$$

as the value of b. This is an integer, since $m \equiv m_1(\bmod n)$, and the equation

$$m + \left(\frac{-m + m_1}{n}\right)n = m_1$$

shows that the value of b is correct.

Finally, we establish the inductive step. Consider a case in which the algorithm requires $s \geqslant 3$ steps and suppose the algorithm is known to give the correct result and only integer values throughout the calculation whenever it ends in fewer than s steps. The induction will proceed from $s - 2$ steps to s steps.



Since $s \geqslant 3$, m_1 and n_1 are nonzero. The first $s - 2$ steps of the calculation result in an integer b_1 such that

$$a_1 m_1 + b_1 n_1 = d$$

for some integer a_1. (Note that the GCD of m_1 and n_1 is the same as that of m and n.) Also, all values calculated up to that point are integers. The entire calculation gives

$$b_0 = b_1 n - d) \div m_1) \times m) + d) \div n.$$

The congruence

$$b_1 n - d \equiv b_1 n_1 - d = -a_1 m_1 \equiv 0 \pmod{m_1},$$

shows that

$$r = (b_1 n - d) \div m_1$$

is an integer. Moreover

$$b_0 = (rm + d) \div n$$

is an integer since

$$rm + d \equiv rm_1 + d = b_1 n \equiv 0 \pmod{n}.$$

Finally, the equation

$$-rm + b_0 n = d$$

shows that $b_0 = b$. The induction is now complete.

It is useful to note that $a = -r$ in the notation above and that r is the number calculated by the first $s - 1$ steps of the algorithm.

Another helpful shortcut is the following: Whenever the algorithm requires at least two steps, the calculation can begin directly to the left of the GCD on the path. In Case 1, it begins at n_{k-1} and the initial value is taken to be n_{k-1}. Thus, in the first example given above, the calculation could begin with

$$10 - 1) \div 9) \times 29) \ldots .$$

In Case 2, the calculation starts at m_{k-1}, but the initial value must be taken to be $-m_{k-1}$. In the second example we would calculate

$$-56 + 7) \div 49) \times 329) \ldots .$$

It is clear that the algorithm described above can be adapted to any Euclidean domain, such as the ring of polynomials over a field.

The algorithm was introduced by the author in a third-year-level course in number theory at California State Polytechnic University, Pomona. As expected, the students quickly became adept at the calculator implementation of the algorithm.

References

1. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, 4th ed., Clarendon Press, Oxford, 1960.
2. W. J. LeVeque, Elementary Theory of Numbers, Addison-Wesley, Reading, Mass., 1962.
3. I. Niven and H. Zuckerman, An Introduction to the Theory of Numbers, 4th ed. Wiley, New York, 1980.

CONSTRUCTING LOOPS FROM GROUPS

ARTHUR H. COPELAND, JR., AND ALBERT O. SHAR
Department of Mathematics & Computer Science, University of New Hampshire, Durham, NH 03824

In any course in group theory it is convenient to have examples of algebraic systems in which most, but not all, of the group postulates are satisfied. For certain kinds of groups $(G, \cdot)$ one can define a new operation "$*$" in terms of "$\cdot$" so that the resulting binary system $(G, *)$ retains all of the defining properties of a group except associativity. Although there are several ways of doing this, the authors have encountered an especially interesting one while studying H-spaces.

If $(G, \cdot)$ is a connected topological group with multiplication $\cdot$ then a new multiplication $*$ may

References

1. R. H. Bruck, A Survey of Binary Systems, Springer-Verlag, 1958.
2. I. M. James, On H-spaces and their homotopy groups, Quart. J. Math. Oxford Ser. (2), 11 (1960) 161–179.

NEW ANGLES ON AN OLD GAME

G. J. PORTER

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104

A standard calculus exercise shows that the trajectory of a projectile, subject only to gravity, is a parabola. If we think of the projectile as a basketball, two additional interesting questions can be asked.

1. What is the smallest initial angle at which a basket can be made?
2. What is the minimum initial velocity needed to make a basket and what angle corresponds to this velocity?

The basketball problem differs from the problem of a bullet hitting a target in one significant detail. The bullet can hit the target either while the bullet is rising or while it is falling. The basketball must go through the hoop while the ball is falling.

We assume that both the ball and the basket are points and that the motion is two-dimensional. The ball is shot from $(0,0)$ at time $t = 0$ with an initial velocity, v_0, and angle, θ. The basket is located at (l, h), where both l and h are positive.

As usual, one easily derives that, at time t, the position of the ball is given by

$$x(t) = v_0(\cos\theta)t \qquad y(t) = -\tfrac{1}{2}gt^2 + v_0(\sin\theta)t$$

where g is the gravitational constant.

A basket is made if there is t_0 such that

(1) $x(t_0) = l$ and $y(t_0) = h$, and

(2) $y'(t_0) < 0$.

Upon our using the equations of motion, these conditions become, respectively,

$$v_0^2 = \left(\tfrac{1}{2}gl^2\sec^2\theta\right)/\,(l\tan\theta - h) \tag{1}$$

and

$$v_0^2 < gl/\,(\sin\theta\cos\theta), \tag{2}$$

which together imply:

THEOREM 1. *For a basket to be made, the initial angle must be greater than* $\mathrm{Tan}^{-1}(2h/l)$.

By substituting $\tan^2\theta + 1$ for $\sec^2\theta$, equation (1) can be transformed into an equation quadratic in $\tan\theta$. This equation has a real solution for $\tan\theta$ when

$$v_0^4 - 2hgv_0^2 - g^2l^2 \geqslant 0.$$

This holds when

$$v_0^2 \geqslant g\left(h + (h^2 + l^2)^{1/2}\right). \tag{3}$$

Thus the minimal initial velocity at which a basket can be made (i.e., for which $\tan\theta$ has a real solution) must satisfy equality in (3). At this velocity

$$\tan\theta = \left(h + (h^2 + l^2)^{1/2}\right)/l. \tag{4}$$

THEOREM 2. *To make a basket with the minimal initial velocity, the initial angle should be*

$$\tfrac{1}{2}(\mathrm{Tan}^{-1}(h/l) + (\pi/2))$$

(i.e., the minimal velocity is required at the angle halfway between the line-of-sight and the vertical).

Proof. We first note that

$$\left(h + (h^2 + l^2)^{1/2}\right)/l > 2h/l;$$

so, by Theorem 1, the basket can be made at the angle θ of equation (4). It thus suffices to show that

$$\tan\left(\tfrac{1}{2}(\mathrm{Tan}^{-1}(h/l) + (\pi/2))\right) = \left(h + (h^2 + l^2)^{1/2}\right)/l.$$

This follows immediately from Fig. 1 by noting that $|AC| = |DC|$ and therefore $\angle DAC = \angle DAE$.

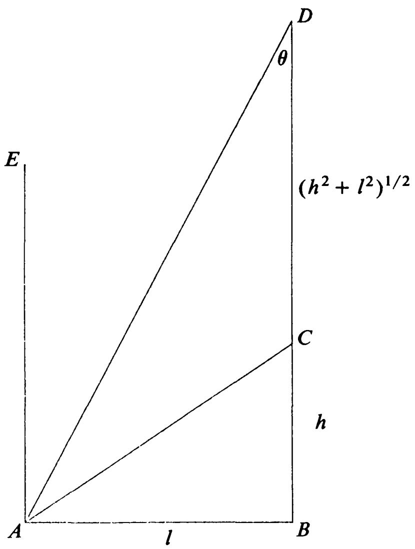


FIG. 1

MATHEMATICAL EDUCATION

EDITED BY MARY R. WARDROP AND ROBERT F. WARDROP

Material for this department should be sent to Robert F. Wardrop, Department of Mathematics, Central Michigan University, Mount Pleasant, MI 48859.

TRANSLATION DIFFICULTIES IN LEARNING MATHEMATICS

JOHN CLEMENT
Department of Physics and Astronomy, University of Massachusetts, Amherst, MA 01003

JACK LOCHHEAD
Department of Physics and Astronomy, University of Massachusetts, Amherst, MA 01003

GEORGE S. MONK
Department of Mathematics, University of Washington, Seattle, WA 98195

Recent mathematics textbooks have increasingly emphasized applications. Mathematical modeling is a critical component of applications, as Rubin [8] points out. Unfortunately, results

of written tests and of interviews suggest that the modeling process is far more complex than is generally imagined. In fact, rather than being an immediate aid to learning mathematics, the process of "translation" between a practical situation and mathematical notation presents the student with a fresh difficulty that must be overcome if the application (or the mathematics) is to make any sense to the student in the long run.

We became aware of this problem during a series of videotaped interviews **[1]** in which college science students were asked to talk aloud while working on simple problems. Using these interviews, we developed a set of written questions (shown in Table 1) that were put to freshman engineering majors from two universities, most of whom were taking calculus. The fact that fewer than 50 percent of these students could solve the problems indicates the difficulty of translating into and out of algebraic notation. The predominant error on the second two problems was reversing the variables in an equation, e.g., $4C = 5S$ instead of $5C = 4S$ in problem 2 in Table 1. The presence of such a consistent pattern suggests that the difficulty is not simply one of misunderstanding English. Furthermore, errors are also high for translations from pictures and data tables **[2]**.

TABLE 1

	# of students tested	% correct
1. Write an equation of the form $P_A =$ ______ for the price you should charge adults to ride your ferry in order to take in an average of D dollars on each trip. You have the following information: Your customers average 1 child for every 2 adults. Childrens' tickets are half-price. Your average load is L people (adults and children). Write your equation for P_A in terms of the variables D and L only.	497	12
2. Write an equation using the variables C and S to represent the following statement: At Mindy's restaurant, for every four people who ordered cheesecake, there were five who ordered strudel. Let C represent the number of cheesecakes ordered and let S represent the number of strudels ordered.	497 most common error: 4C = 5S	39
3. Weights are hung on the end of a spring and the stretch of the spring is measured. The data are shown in the table below: Stretch S (cm) — Weight W (g) 3 — 100 6 — 200 9 — 300 12 — 400 Write an equation that will allow you to predict the stretch (S) given the weight (W).	381 most common error: 3S = 100W	42

To make certain that our results were not the consequence merely of inattention, one of us included a problem similar to problem 2 as part of a final examination in calculus. More than 40 percent failed this problem.

Sources of Error—The Reversal. At first, the students' difficulty in translation greatly

surprised us. We had not been looking for it and, in fact, had assumed that college students could translate between English and algebra, at least in simple situations. After discovering the difficulties, however, we recalled that students are rarely asked to *construct* a formula or to interpret one in a significant way. They are usually given a formula or asked to select the appropriate formula from a well-defined (and very short) list and then to manipulate it using techniques from algebra or calculus. The one place in the secondary school mathematics curriculum where translation plays a large role is in doing "story problems." But we suspect that teachers have tended to deemphasize such problems because students find them difficult.

To investigate the source of the errors we had observed, we collected data on the following simpler problem:

The Students-and-Professors Problem. Write an equation for the following statement: "There are six times as many students as professors at this university." Use S for the number of students and P for the number of professors.

On a written test with 150 calculus-level students, 37 percent missed this problem, and two-thirds of the errors took the form of a reversal of variables such as $6S = P$. In a sample of 47 nonscience majors taking college algebra, the error rate was 57 percent.

We also interviewed 15 students who were asked to think out loud while solving problems like this one. The videotaped records provide a much more detailed account of the students' thought processes than is possible with written tests. They allow one to distinguish between insignificant careless errors and more serious conceptual problems. Several interviews on problems like this one lasted more than 5 minutes. In these interviews the students vacillated between correct and incorrect solutions and appeared to be thoroughly confused, not just guilty of making hasty mistakes.

By analyzing the transcripts we identified two distinct sources for the students' tendency to reverse variables. The first faulty approach, which we call "word order matching," is described by Paige and Simon as "syntactic translation" [**5**]. This is a literal, direct mapping of the words of English into the symbols of algebra. For example, one might make the translation:

There are 6 times as many students as professors

$$6 \cdot S = P$$

Here the direct mapping has led to a reversed solution. We note that some textbooks explicitly instruct students to translate by the syntactic method [**3**].

We call the second method of mistranslation the "static-comparison" method. The student who takes this approach understands that the sentence implies that the student population is much bigger than the faculty population; in some cases the student will draw a diagram to indicate that this is so. (See Fig. 1.) But the student still believes that this relationship should be

FIG. 1

represented by the equation $6S = P$. Apparently the expression "$6S$" is used to indicate the larger group and "P" to indicate the smaller group. The letter S is not understood as a variable that represents the *number* of students but rather is treated like a label or unit attached to the number 6. The equals sign expresses a comparison or association, not a precise equivalence. This interpretation of the equation is a literal attempt to symbolize the static comparison between two groups. This approach was used frequently by the students we interviewed. It may seem "more wrong" than the syntactic approach, but it does have the virtue of starting from a representation

of the essential features of the problem. We regard this as an important first step in the translation process.

People who did a wide variety of problems correctly used a markedly different approach, which we call "operative translation." This approach requires the comprehension of the static-comparative approach, together with a much richer sense of what a mathematical equation is and says. In the students-and-professors problem, the number S is seen as bigger than P; therefore, the number P must be operated on by multiplication by 6 to produce a number that is the same as S. This is a lot to squeeze into the sentence $6 \cdot P = S$, but it is exactly what is required in order to understand the meaning of the simplest algebraic equations.

The reversal difficulty appears to be rather resilient and requires considerable attention and discussion before students can learn to overcome it. In a study of calculus students (primarily engineering and science majors) Rosnick and Clement [6] report on the effects of a 15–30 minute teaching unit involving worked examples and practice with several problems of this kind. Tapes of the individual teaching sessions showed that most of the students were not able to develop a reliable understanding of the issue after a moderate amount of tutoring.

Conclusion. The reversal error appears to be due not simply to carelessness but rather to a self-generated, stable, and persistent misconception concerning the meaning of variables and equations. The concepts of variable and equation are so fundamental that it is hard for practiced users of mathematics to imagine how such misconceptions can persist. However, it is not surprising that this misconception has not been affected by years of practice with manipulation of equations, since these techniques usually do not require one to understand the meaning of an equation.

Even after taking a semester or more of calculus, many students have difficulty expressing relationships algebraically. They cannot translate reliably between algebra and other symbol systems, such as English, data tables, and pictures. We do not believe that this is a trivial problem. Apparently it is rare for mathematicians to think solely in terms of algebraic symbols [4], [7], [9]. Rather, they often describe their thoughts as being like pictures, diagrams, or graphs. At some point, the mathematician is able to translate these ideas into algebraic notation; this translation is precisely what our students have not learned to do! The outlook is just as bleak for those who will never make an original mathematical contribution. They must learn to apply mathematics; that is, they must translate a problem usually expressed in words into algebraic notation and retranslate a solution back into words. Thus, translation skills are critically important in learning mathematics.

What makes teaching (and learning) of these translation skills so difficult is that behind them there are many unarticulated mental processes that guide one in constructing a new equation on paper. These processes are not identical with the symbols; in fact, the symbols themselves, as they appear on the blackboard or in a book, communicate to the student very little about the processes used to produce them. There seems to be no way to explain such translation processes to students quickly.

Our own experience suggests that one method for helping students to acquire these skills is to: (1) allocate time in courses for developing and practicing them as separate skills; (2) assign translation problems (such as those discussed above) that cannot be solved by trivial syntactic or other nonoperative approaches; (3) show by many examples the shortcomings of the latter methods; and (4) emphasize the operative nature of equations. These techniques have given us encouraging preliminary results. We are continuing our investigations into the best methods for teaching this important skill. We hope that some readers may be interested in trying experiments with their own students and that they will join us in investigating translation skills, a long-neglected component of mathematical literacy.

The authors thank P. Rosnick and R. Narode for help in collecting these data, and M. Janowitz and F. Byron for comments on an earlier version of this paper.

This research was supported in part by NSF award No. SED78-22043 in the Joint National Institute of Education—National Science Foundation Program of Research on Cognitive Processes and the Structure of Knowledge in Science and Mathematics.

References

1. J. Clement, Algebra word problem solutions: An analysis of a common misconception, paper presented at a meeting of the American Educational Research Association, Boston, 1980.

2. J. Clement, J. Lochhead, and E. Soloway, Translating between symbol systems: Isolating a common difficulty in solving algebra word problems (technical report), Department of Physics and Astronomy, University of Massachusetts at Amherst, 1979.

3. M. Dolciani et al., Modern School Mathematics: Algebra and Trigonometry II, Houghton Mifflin, Boston, 1968.

4. J. Hadamard, The Psychology of Invention in the Mathematical Field, Princeton University Press, Princeton, N.J., 1945.

5. J. Paige and H. Simon, Cognitive processes in solving algebra word problems, in B. Kleinmuntz, ed., Problem Solving Research, Method and Theory, John Wiley, New York, 1966.

6. P. Rosnick and J. Clement, Learning without understanding: The effect of tutoring strategies on algebra misconceptions, Journal of Mathematical Behavior, 3 (1980), in press.

7. G. Pólya, Mathematics and Plausible Reasoning, vol. 1: Induction and Analogy in Mathematics, Princeton University Press, Princeton, N.J., 1954.

8. R. Rubin, Model formulation using intermediate systems, this MONTHLY, 86 (1979) 299–303.

9. R. Shepard, The mental image, American Psychologist (1978) 125–137.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA (USA) 94303, by August 31, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2878. *Proposed by M. Slater, University of Bristol, U.K.*

When does a real quartic have no real zeros?

E 2879. *Proposed by Paul W. Haggard, East Carolina University.*

Let D be an integral domain with characteristic p (prime). The expansion of $(x+y)^n$ (in $D[x,y]$) has $N = \Pi_{i=0}^{m}(k_i+1)$ nonzero terms, where $n = \Sigma_{i=0}^{m} k_i p^i$ is the p-expanded form of the positive integer n. Clearly, $N \leqslant n+1$. (a) For what integers n is $N = n+1$? (b) Can the relation $N = n$ (>0) hold?

E 2880. *Proposed by Leo J. Alex, SUNY, College at Oneonta.*

Find all solutions to the equation $x^2 + 15^a = 2^b$ in integers x, a, and b.

E 2881. *Proposed by Nick Franceschine III, Sebastopol, Calif.*

Enumerate the positive rationals as Cantor did, $r_1 = 1$, $r_2 = 1/2$, $r_3 = 2$, $r_4 = 3$, $r_5 = 1/3$, $r_6 = 1/4$, $r_7 = 2/3, \ldots$. Exhibit an infinite set of real numbers outside every one of the closed intervals $[r_n - 2^{-n-1}, r_n + 2^{-n-1}]$.

E 2882. *Proposed by Ronald J. Evans, University of California at San Diego*

Let p be a prime congruent to any one of $5, 7, 11, 13, 23$ modulo 24. Write (n/p) for the Legendre symbol. Show that $\Sigma n(n/p) = 0$, if the sum is extended over all values of n congruent to 1 mod 6 in the range $0 < n < 6p$.

E 2883. *Proposed by J. M. Patin, St. Herblain, France.*

Can integers a, b, c be found so that the nonconstant polynomial $an^2 + bn + c$ has all its prime factors congruent to 3 (mod 4) for $n = 1, 2, \ldots$?

SOLUTIONS OF ELEMENTARY PROBLEMS

Orthogonal Triad Meeting a Conic

E 2751 [1979, 56]. *Proposed by Paul Monsky, Brandeis University.*

Let X be a conic section. Through what points in space do there pass three mutually perpendicular lines, all meeting X?

Solution, adapted from separate solutions by Theodore S. Bolis, SUNY at Oneonta, and R. K. Oliver, Pittsburgh, Pa. Suppose the conic is given by $H(x,y) = 0$ where $H(u_1,u_2) = au_1^2 + bu_1u_2 + cu_2^2 + du_1 + eu_2 + f$, with a, b and c not all zero. Let S be the quadric surface $(a+c)z^2 + H(x,y) = 0$. Then the locus is contained in S and is all of S except in one case—when the conic consists of two perpendicular lines. In particular, when the conic is nondegenerate the locus is either an ellipsoid, a paraboloid of revolution, a hyperboloid, or an equilateral hyperbolic cylinder.

Suppose that $P_0 = (x_0,y_0,z_0)$ is in the locus with $z_0 \neq 0$. Let $P = (x,y,z)$ be a unit vector with $z \neq 0$. The line $P_0 + tP$ meets the xy-plane at $P^* = z^{-1}(x_0z - z_0x, y_0z - z_0y)$. P^* lies on the conic precisely when $G(x,y,z) = a(x_0z - z_0x)^2 + b(x_0z - z_0x)(y_0z - z_0y) + c(y_0z - z_0y)^2 + dz(x_0z - z_0x) + ez(y_0z - z_0y) + fz^2 = 0$. G is a homogeneous quadratic polynomial, and the coefficients of x^2, y^2, and z^2 in G are az_0^2, cz_0^2 and $H(x_0,y_0)$, respectively. Since there are three mutually perpendicular lines through P_0 meeting the conic, there are three mutually orthogonal unit vectors $P = (x,y,z)$, each of which satisfies the condition $G(P) = G(x,y,z) = 0$. Thus the symmetric matrix A attached to G is orthogonally similar to a matrix with zeros as diagonal entries, trace $A = 0$, $(a+c)z_0^2 + H(x_0,y_0) = 0$, and P_0 lies in S.

Conversely suppose that P_0 is in S with $z_0 \neq 0$. Let $Q_i : (x,y,0)$ be the unit vectors in the xy-plane with $ax^2 + bxy + cy^2 = 0$. (There will be 0, 2, or 4 of them.) Since $G(x,y,0) = z_0^2(ax^2 + bxy + cy^2)$, the Q_i are precisely the unit vectors Q in the xy-plane with $G(Q) = 0$. Now since $P_0 \in S$, trace $A = 0$, A is indefinite, and we can find a unit vector P with $G(P) = 0$.

If we assume that G has no linear factor of the form $\alpha x + \beta y$, we may even choose P so that it is not equal to or orthogonal to any Q_i. Since trace $A = 0$ we can complete P to an orthonormal basis P, P', P'' with $G(P) = G(P') = G(P'') = 0$. Now none of P, P' or P'' is equal to any Q_i. It follows that they do not lie in the xy-plane. We then get three mutually perpendicular lines through P_0 meeting the conic. Suppose now that $a + c \neq 0$. Then the coefficient of z^2 in G is $-(a + c)z_0^2 \neq 0$, G has no factor of the form $\alpha x + \beta y$, and we may argue as above. If $a + c = 0$ we may assume by a change of coordinates that $H(x,y) = xy + f$. Suppose $f \neq 0$. Then

$$G(x,y,z) = (x_0 z - z_0 x)(y_0 z - z_0 y) + fz^2 = z_0(z_0 xy - y_0 xz - x_0 yz).$$

Since $x_0 y_0 = -f \neq 0$, G is irreducible, has no factor of the form $\alpha x + \beta y$, and we're done. The case of two perpendicular lines is indeed exceptional. Those points of S lying over the intersection are not in the locus.

Theodore Bolis (and the proposer) generalized this result to higher dimensions. Let C be the quadratic $(n-1)$-fold in R^{n+1} defined by the equations $H(x_1,\dots,x_n) = 0$, $x_{n+1} = 0$, where H has total degree 2. Then those points of R^{n+1} through which there pass $n+1$ mutually perpendicular lines all meeting C form a quadratic n-fold defined by the equation $(\Sigma a_{ii})\cdot x_{n+1}^2 + H(x_1,\dots,x_n) = 0$ where a_{ii} is the coefficient of x_i^2 in H. When $n \geqslant 3$ the exceptional case described above does not arise.

Also solved by the proposer.

Linear Transformation of $\dot{x} = Ax$

E 2814 [1980, 60]. *Proposed by Walter Leighton, University of Missouri, Columbia.*

Consider the system

$$\dot{x} = ax + by, \qquad \dot{y} = cx + dy, \tag{1}$$

where a, b, c, d are real constants with $(a-d)^2 + 4bc < 0$. It is well known that there exist linear affine transformations

$$x = pu + qv, \qquad y = ru + sv \qquad (p,q,r,s \text{ real constants})$$

of (1) such that

$$\dot{u} = hu - kv, \qquad \dot{v} = ku + hv. \tag{2}$$

Find one such set of values p, q, r, s.

Solution. Set

$$X = (x,y)^*, \quad U = (u,v)^*, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad H = \begin{pmatrix} h, & -k \\ k, & h \end{pmatrix}.$$

The problem asks for a real matrix $P = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ such that $AP = PH$. One solution is $p = -2b$, $q = 0$, $r = a - d$, $s = -(-r^2 - 4bc)^{\frac{1}{2}}$.

The solvers were D. W. Bailey, K. L. Bernstein, W. Boucher (Canada), R. Breusch, H. Carus, F. S. Cater, R. J. Evans, N. Franceschine III, J. R. Hatcher, E. Heil (Germany), B. R. Hunt (student), A. A. Jagers (Netherlands), H. Kappus (Switzerland), L. Kuipers (Switzerland), R. T. Lee (Canada), G. N. Lewis, D. F. Lockhart, O. P. Lossers (Netherlands), L. E. Mattics, B. J. McCartin, R. K. Oliver, A. Shuchat, D. A. Singer, St. Olaf Problem Group, W. V. Webb, J. Wiener, and the proposer.

Bernstein, Kuipers, Lewis, Lockhart, found all solutions P. The matrix H is uniquely determined by A.

The Inequality $\cos x + \cos y \leqslant 1 + \cos xy$

E 2815 [1980, 136]. *Proposed by Leon Gerber, St. John's University, Jamaica, N.Y.*

Establish the inequality

$$\cos x + \cos y \leqslant 1 + \cos xy$$

for $0 \leqslant x^2+y^2 \leqslant \pi$.

I. *Solution by Brian R. Hunt (sophomore), Montgomery Blair High School, Silver Spring, Md.* Since cosine is an even function, we may assume $x, y \geqslant 0$. Furthermore, we may assume $x, y > 1$ since if $x \leqslant 1$, then $xy \leqslant y \leqslant \sqrt{\pi} < \pi$, $\cos y \leqslant \cos xy$, and hence $\cos x + \cos y \leqslant 1 + \cos xy$. If the inequality is false, then it follows from $xy \leqslant (x^2+y^2)/2 \leqslant \pi/2$ that $\cos xy > 0$ and hence $\cos x + \cos y \geqslant 1$. Consequently, $\cos x \geqslant 1 - \cos y > 1 - \cos 1 > .45$ and $x < \cos^{-1} .45 < 1.2$. Similarly, $y < 1.2$ and $xy < (1.2)^2 = 1.44$. Therefore, $1 + \cos xy > 1 + \cos 1.44 > 1.13 > 2\cos 1 > \cos x + \cos y$, a contradiction.

II. *Solution by Theodore S. Bolis, State University College at Oneonta, New York.* Since $1 - x^2/2 \leqslant \cos x \leqslant 1 - x^2/2 + x^4/24$ for all real x, it is enough to establish that

$$1 - x^2/2 + x^4/24 + 1 - y^2/2 + y^4/24 \leqslant 1 + 1 - x^2y^2/2,$$

i.e.,

$$x^4 + y^4 + 12x^2y^2 - 12(x^2+y^2) \leqslant 0 \qquad \text{for } x^2+y^2 \leqslant \pi. \tag{1}$$

In polar coordinates, (1) is equivalent to

$$r^2(2 + 5\sin^2 2\theta) \leqslant 24 \qquad \text{for } r^2 \leqslant \pi.$$

Since

$$r^2 \leqslant \pi \text{ and } \sin^2 2\theta \leqslant 1,$$

this is true if $7\pi < 24$, which is correct.

Also solved by Miroslav D. Ašić (Yugoslavia), Ben B. Bowen, Erhard Braune (West Germany), Robert Breusch, J. L. Brown, Jr., F. S. Cater, Chico Problem Group, Roger Cuculière (France), Peter Z. Daffer, G. Fourt (France), P. Eenigenburg & A. Stroddart, Robert Gilmer, Noel Glick, E. Grosswald, Joel K. Haack, Sidney Heller, A. A. Jagers (Netherlands), L. Kuipers (Switzerland), O. P. Lossers (Netherlands), Aries Matsoukas & Achilles Venetoulias (Greece), D. R. Morrison, M. P. Ojha, Roger B. Nelson, Eero Posti (Finland), Otto G. Ruehr, Michael Skalsky, Abraham Smuckler (Israel), R. S. Stacy (West Germany), J. Suck (West Germany), Daniel Weisser, Gerald Wildenberg, and the proposer.

Grosswald showed that the inequality remains valid for $x^2+y^2 \leqslant 4.1$.

A Function Involving $\Sigma\sigma(k)$

E 2817 [1980, 137]. *Proposed by Jeffrey Shallit, undergraduate, Princeton University.*

For k in $\{1, 2, \ldots, n\}$, let $R(k,n)$ be the remainder in the division of n by k. Thus $n = qk + R(k,n)$ with q an integer and $0 \leqslant R(k,n) < k$. Set $T(n) = \Sigma_{k=1}^{n} R(k,n)$.

(i) Express $T(n)$ in terms of the function σ, where $\sigma(n)$ is the sum of the positive integral divisors of n.

(ii) Show that $T(2^n) = T(2^n - 1)$.

Solution. We use $\Sigma k[n/k] = \Sigma\sigma(k)$, the summation being extended from $k = 1$ to $k = n$. Then $T(n) = \Sigma(n - qk) = n^2 - \Sigma\sigma(k)$. Since $\sigma(2^n) = 2n - 1$, the assertion (ii) follows. [Alternatively, the formula for $T(n)$ can be obtained by induction, starting with the recursion $T(n) = T(n-1) + (2n-1) - \sigma(n)$; (ii) is a special case of this recursion.]

The solvers were H. L. Abbott (Canada), B. B. Bowen, K. Brown, R. Breusch, P. S. Bruckman, F. S. Cater, G. Fenstamaker, M. Johnson, E. Tarkonish & B. Leonard (jointly), N. Franceschine, A. L. Furno, R. Heller, M. F. Kruelle, K. Y. Li, R. B. McNeill, J. L. deMiguel (Spain), V. S. Ramaiah (India), H. J. Ricardo, J. P. Robertson, University of South Alabama Problem Group, L. van Hamme (Belgium), B. Viswanathan (Canada), C. R. Wall, P. Zwier, and the proposer.

Application of the Euler-Fermat Theorem

E 2818 [1980, 137]. *Proposed by J. Linkovskiĭ-Condé, Moscow, U.S.S.R.*

Let $T_n = 2^n + 1$ for all positive integers. Let ϕ be the Euler totient function, let k be any

positive integer, and $m = n + k\phi(T_n)$. Show that T_m is divisible by T_n.

Solution by Bela Brindza, student, Debrecen, Hungary. By the Euler-Fermat theorem, $2^{\phi(T_n)} \equiv 1 \bmod T_n$. Since $T_m = 2^n 2^{k\phi(T_n)} + 1$, it follows that $T_m \equiv 2^n + 1 \bmod T_n$.

Also solved by 52 other readers, including the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by August 31, 1981. The solver's full post-office address should be on each sheet.

6339. *Proposed by P. Erdős, Hungarian Academy of Sciences, and J. L. Selfridge, Mathematical Reviews.*

Let $\Pi_1^s p_i^{\alpha(i)}$ be the canonical (prime) factorization of $n!$, so that $\pi(n) = s$. Let $f(n) = j$ be the unique index for which

$$\prod_1^{j-1} p_i^{\alpha(i)} < (n!)^{1/2} < \prod_1^{j} p_i^{\alpha(i)}.$$

Prove

(i) $|f(n+1) - f(n)| \leqslant 1$.

(ii) As n increases, each of the relations

$$\prod_1^{j-1} p_i^{\alpha(i)} \gtrless \prod_{j+1}^{s} p_i^{\alpha(i)},$$

has infinitely many solutions.

(iii) $n^{1/2}/p_j$ approaches a fixed limit.

*(iv) For infinitely many n, $f(n+1) = f(n) - 1$. For all $m > n$, $f(m) \geqslant f(n) - 1$.

6340. *Proposed by J. M. Patin, St. Herblain, France, and H. Stark, University of California, San Diego.*

Given a nonconstant polynomial $f(x)$ over $\mathbb{Z}$ and a positive $m \in \mathbb{Z}$, show that there exists $n \in \mathbb{Z}$ and a prime $p \equiv 1 \pmod m$ such that p divides $f(n)$. (For a special case of this problem see Elementary Problem E 2883.)

6341. *Proposed by J. C. Lagarias, Bell Laboratories, and H. W. Lenstra, University of Amsterdam.*

Let L be a line in the complex plane $\mathbb{C}$ passing through two relatively prime algebraic integers. (Two algebraic integers α, β are *relatively prime* if there exist algebraic integers γ, δ such that $\alpha\delta + \beta\gamma = 1$.) Prove that there are infinitely many algebraic units lying on the line L.

6342. *Proposed by Richard Stanley, Massachusetts Institute of Technology.*

Let $f(n)$ be the number of nonisomorphic n-element partially ordered sets P which do not contain three pairwise incomparable elements. (Equivalently, P is a union of two chains.) Let $F(x) = 1 + \Sigma_{n>1} f(n)x^n = 1 + x + 2x^2 + 4x^3 + 10x^4 + \cdots$. Show that

$$F(x) = \frac{4}{2 - 2x + \sqrt{1-4x} + \sqrt{1-4x^2}}.$$

6343. *Proposed by W. Holsztynski, Ann Arbor, Michigan.*

(a) Show that there exists a four element metric space which cannot be isometrically embedded in a Hilbert space.

(b) Let X be the Banach space obtained from $\mathbb{R}^2$ by defining the norm $\|(x,y)\| = \max\{|x|,|y|\}$. Show that every four-element metric space can be isometrically embedded in X.

SOLUTIONS OF ADVANCED PROBLEMS

Preservation of Convexity Under Multiplication

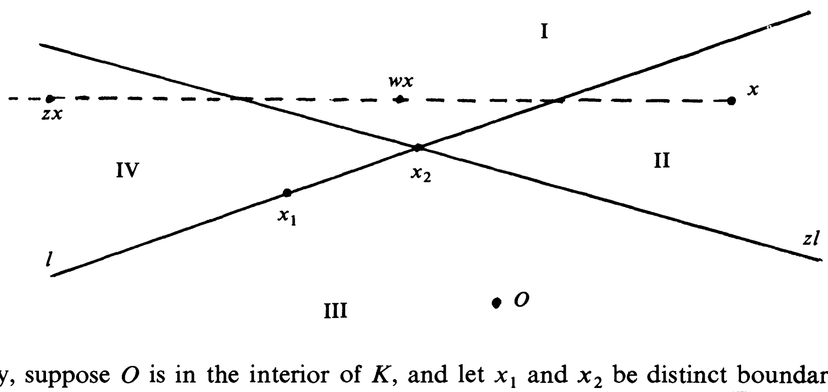
5297[1973, 814]. *Proposed by C. R. Johnson, National Bureau of Standards.*

Find all convex subsets K of the complex plane C such that if L is any convex subset of C, then $\{zw\colon z \in K, w \in L\}$ is convex.

Solution by Robert Israel, University of British Columbia. K is either a single point, the whole plane C, a ray from O or interval thereof, or a disk centered at O with any subset of its boundary. It is easy to see that all these cases satisfy the hypothesis.

If O is not in the interior of K, suppose K subtends an angle $\alpha > 0$ at O (i.e., the smallest sector containing K has opening angle α). Let L be any convex set subtending an angle β at O with $\pi < \alpha + \beta < 2\pi$. Then $LK \equiv \{zw\colon z \in L, w \in K\}$ subtends an angle $\alpha + \beta$, which is impossible for a convex set. So if O is not in the interior of K, then K is contained in a ray through O.

Now suppose O is in the interior of K. I claim the boundary of K contains no straight line segments. Otherwise, let x_1 and x_2 be distinct points in the interior of such a segment on the line l, and let L be the line segment from 1 to $z \equiv x_2/x_1$ (see figure). Since LK subtends an angle greater than π at x_2, it suffices to show x_2 is not in the interior of LK. In fact, suppose wx is in region (I) with $w \in L$ and $x \in K$. Then wx is a convex combination of x and zx. Since K (respectively, zK) is contained in regions (II) and (III) (respectively, (III) and (IV)), we must have x in (II) and zx in (IV). But this is impossible, since (with the orientation as in the figure) the argument of z is between $-\pi$ and 0. This proves the claim.



Finally, suppose O is in the interior of K, and let x_1 and x_2 be distinct boundary points, l_1 and l_2 lines through x_1 and x_2, respectively, containing no other points of $\bar{K}$ (these exist by the Hahn-Banach Theorem and the paragraph above). There is some $z \in C$ such that $zl_2 = l_1$. Let L be the line segment from 1 to z. By hypothesis, $\frac{1}{2}x_1 + \frac{1}{2}zx_2 \in \overline{(LK)} = L\bar{K}$ (using compactness of L), and thus is a convex combination of x and zx for some $x \in \bar{K}$. But $\bar{K}$ and $z\bar{K}$ are on the same side of l_1, intersecting l_1 only at x_1 and zx_2, respectively, so if $x_1 \neq zx_2$, $\frac{1}{2}x_1 + \frac{1}{2}zx_2$ cannot be written as a convex combination of elements of $\bar{K}$ and $z\bar{K}$ in any other way. But then $x_1 = x = x_2$, which is a contradiction. Therefore we must have $x_1 = zx_2$. This means that the angle between l_1 and the line from O to x_1 equals the angle between l_2 and the line from O to x_2. In particular, the tangent to K at each boundary point is unique and makes a constant angle with the radius. The only curves with this property are logarithmic spirals and circles centered at O. Since a spiral is not the boundary of a convex set, K must be a disk centered at O (with any subset of its boundary).

Derivatives of Continuous Functions

6140* [1977, 221]. *Proposed by F. S. Cater, Portland State College, Oregon.*

Let f be a continuous real-valued function on $[0, 1]$ and let E_f denote the (possibly void) set $\{x \in [0,1]\colon f'(x) \text{ exists and is finite}\}$. Let $a(f)$ = Lebesgue outer measure of $f([0,1]\setminus E_f)$, and let

$$m(t) = \begin{cases} f'(t) & \text{for } t \in E_f \\ 0 & \text{otherwise.} \end{cases}$$

Let $b(f) = a(f) + \int_0^1 m(t)\,dt$ and $c(f) = a(f) + \int_0^1 [1 + m(t)^2]^{1/2}\,dt$.

Find (1) $\max c(f)$ over all f such that $b(f) = 1$, and (2) $\min c(f)$ over all f such that $b(f) = 1$. Describe functions for which $c(f)$ takes one of these values.

Solution by K. F. Andersen, University of Alberta, Canada. By Minkowski's inequality for integrals [see, for example, Hardy, Littlewood, and Pólya, *Inequalities*, p. 146],

$$\begin{aligned} c(f) &= a(f) + \int_0^1 \left[1 + m(t)^2\right]^{1/2} dt \\ &\geqslant a(f) + \left\{\left(\int_0^1 1\,dt\right)^2 + \left(\int_0^1 m(t)\right)^2\right\}^{1/2} \end{aligned} \tag{1}$$

with equality if and only if $m(t) = k$ a.e., k a positive constant. Now $b(f) = 1$ yields

$$c(f) \geqslant a(f) + \left\{1 + [1 - a(f)]^2\right\}^{1/2} \tag{2}$$

and, for $0 \leqslant a(f) \leqslant 1$, the term on the right-hand side of (2) is increasing in $a(f)$; hence

$$c(f) \geqslant \sqrt{2}\,, \tag{3}$$

with equality in (3) if and only if $a(f) = 0$ and $m(t) = k$ a.e.; moreover, $k = 1$ in this case since $b(f) = 1$. Thus $c(f)$ is minimal for functions of the form $f(t) = t + g(t)$ where g is continuous, almost everywhere differentiable, and $g'(t) = 0$ a.e. on $[0, 1]$. On the other hand, the requirement $b(f) = 1$ does not guarantee that $c(f)$ is bounded as the example $f_n(t) = \sin(2n + 1/2)\pi t$ clearly shows, since $b(f_n) = 1$ while $c(f_n)$, the arc length of the graph of $f_n(t)$, certainly exceeds $2n$.

Analytic Characterization of Convexity

6166 [1977, 576]. *Proposed by D. A. Gregory, Queens University, Kingston, Ontario.*

If f is a convex functional on a convex subset K of a vector space, then for all x and $x + h$ in K, the one-sided directional derivatives

$$f'_+(x,h) = \lim_{a\to 0^+} \frac{f(x + \alpha h) - f(x)}{\alpha}$$

exist in the extended reals and $f(x + h) \geqslant f(x) + f'_+(x,h)$. Is the converse true? If so, we have an analytic characterization of convex functionals.

Solution by P. D. Taylor and the proposer, Queen's University, Canada. As usual, we may restrict to the case where K is an interval on the real line [**1**, p. 91]. In that setting, the problem observes that if a function is convex, then with the customary one-sided restrictions at endpoints, its (possibly infinite) right and left derivatives exist at each x in K and the graph of the function is on or above the (possibly straight up or straight down) tangent rays to the right and left at each point on the graph. This is well known [**1**, p. 6]; the problem asks if the converse holds.

The converse is false. For a counterexample, let K be the real line and take $f(x) = 1$ if x is an integer, $f(x) = 0$ otherwise. The converse does hold, even for general K, if we also assume that $f'_+(x,h)$ is finite when x is an interior point of a line segment parallel to h in K. This "finite interior slopes" condition is necessary for f to be convex [**1**, p. 5]. It implies that the restriction of

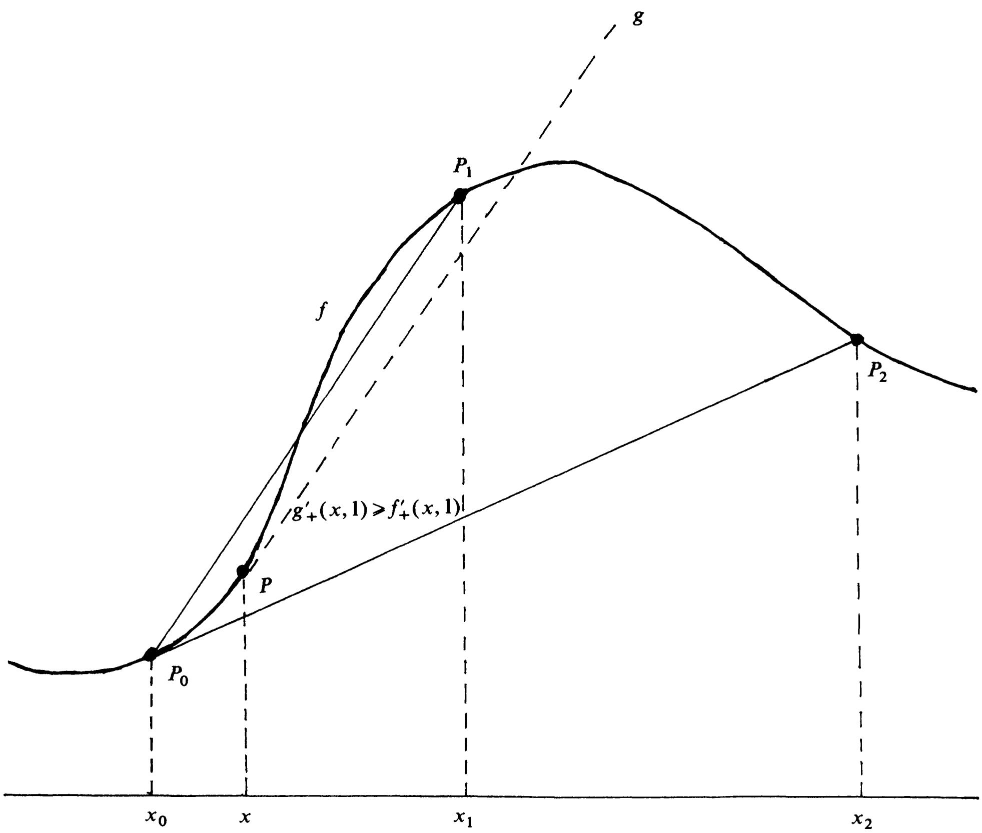


FIG. 1

f to an interval in K is continuous at each interior point x of the interval; for if $x \pm \alpha h$ are in the interval, then $f(x \pm \alpha h) - f(x) \to 0$ as $\alpha \to 0^+$.

We prove the converse assuming the "finite interior slopes" condition. To begin with, f is convex on $\operatorname{int} K$, the interior of the interval K. If not, then there are x_0, x_1, x_2 in $\operatorname{int} K$ such that P_1 is above the chord P_0P_2, where $x_0 < x_1 < x_2$ and $P_i = (x_i, f(x_i))$, $i = 0, 1, 2$ (see Fig. 1). In fact, using the continuity of f, we may select $x_0 < x_2$ so that $P = (x, f(x))$ is above the chord P_0P_2 whenever $x_0 < x < x_2$. The given inequality implies that for $x > x_0$, x in K, $P = (x, f(x))$ is on or above the right tangent ray to the graph of f at P_0. In particular, slope $P_0P_2 \geqslant f'_+(x_0, 1)$, the right derivative of f at x_0. Now slope $P_0P_1 >$ slope P_0P_2 and slope $P_0P \to f'_+(x_0, 1)$ as $x \to x_0^+$. Thus, by the continuity of f on $\operatorname{int} K$, slope $P_0P_2 <$ slope $P_0P <$ slope P_0P_1 for some P with $x_0 < x < x_1$. Moreover, $f'_+(x, 1) \geqslant$ slope P_0P_1 for at least one such P; for example, choose by the continuity of f a point P on the graph of f at the greatest vertical distance below the chord P_0P_1 and note that $(g - f)'_+(x, 1) \geqslant 0$, where g is the right tangent function (this suggests a "mean value theorem" for continuous functions with right derivatives). Now P_2 must be below the right tangent ray to the graph of f at P. This contradicts the given inequality. Thus f must be convex on $\operatorname{int} K$.

If K has a right endpoint b, then we must have $f(b) \geqslant \limsup_{\alpha > 0} f(b - \alpha)$: otherwise $f'_+(b, -1) = \infty$ and the given inequality could not hold for a real valued function f. A similar inequality would hold if f had a left endpoint. It follows that f is convex on the whole interval K.

Reference. A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York, 1973.

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

We invite readers to submit reviews of significant recent college-level mathematics books. We especially encourage reviews based on classroom use and comparative reviews of several related books. Reviews should ordinarily not exceed two pages (per book) typed double spaced. Manuscripts of reviews as well as books submitted for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, MN 55057.

Ordinary Differential Equations with Modern Applications. By N. Finizio and G. Ladas. Wadsworth Publishing, New York, 1978. xv + 380 pp. $15.95. (Telegraphic Review, May 1978.) *Solutions for Ordinary Differential Equations with Modern Applications*, 1978, 260 pp.

Most students' first real experience with applied mathematics is the sophomore-level course in differential equations, although they may have seen some differential equations in physics courses. This text allows them to study a much wider variety of applications than with earlier texts. Unlike the window dressing in some of its direct competitors, the development of realistic applications is integral to this book. In the introduction, the 6-page list of applications ranges from approximation theory, biology, and diffusion, through electromechanical systems, mathematical physics, and mechanics, to learning theory, quantum mechanics, and statistics. In addition, applications are often included in assigned problems taken from journals in biology, psychology, and other areas, as well as from this MONTHLY and *Mathematics Magazine*.

We offer a course, using the first four chapters, that covers the standard topics, reduction-of-order techniques, Laplace transforms, and the solution of linear systems by matrix procedures. Our students really appreciate the power of such techniques as illustrated in examples of electrical circuit systems, interacting species in ecological systems, and economic systems of interrelated markets. Such interesting applications have made it easier to motivate some of the mathematical techniques.

Colleagues tell me that the main difficulty encountered with the text was how to decide which of the many excellent applications to consider. With this text it is almost impossible to teach a "cookbook" course. In fact, on a number of occasions students have initiated discussions on the implications of various models. The large number of examples from business and economics and other nonphysics applications appear to interest the students (and instructor) strongly. Is it possible that some texts on differential equations evoke a form of "physics anxiety"?

The solutions manual has complete solutions for all the exercises. This can be very helpful in assigning problems and designing exam questions. In addition, the complete detail of the solutions makes it an extremely helpful aid to the students, *provided* they first attempt to solve the problems without the manual. Faculty opinion is divided on whether the solutions manual should be available to students. I permitted my students to use the manual and feel that it has resulted in a noticeable increase in both quantity and quality of student work.

The final four chapters consider series solutions of second-order differential equations, boundary value problems, numerical solutions of differential equations, and nonlinear differential equations. This suggests an interesting syllabus for a second course in differential equations.

I have offered the Systems Theory part in a refresher course for engineers preparing to take the Engineering in Training (E.I.T) Fundamentals examination for the licensing of professional engineers. Much of the material covered in this course is considered in the text, including material on feedback, the stability of systems and transform solutions of electrical circuit systems. The text is excellent for engineers reviewing for the exam, because it is very readable. In fact the fine presentation makes it an excellent reference source for both students and faculty.

Students have found the text very appropriate. It has been used for two years in our program and the faculty who have used it have commented on its originality and excellence and want to continue to use it.

RON BARNES, University of Houston-Downtown College

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, S*, P*, L*. It Seems I Am a Jew: A Samizdat Essay. Grigori Freiman. Trans: Melvyn B. Nathanson. So Ill U Pr, 1980, xvi + 96 pp, $9.95. [ISBN: 0-8093-0962-9] A moving, detailed, personal account of the increasing purge of Jews from Russian mathematics. Details of events, letters, "Jewish" problems on entrance exams, and patterns of rejected dissertations are backed up by specific names of those responsible. A strong and damning accusation against the current Soviet mathematics leadership. LAS

General, P*, L*. Encyclopedic Dictionary of Mathematics. Ed: Shôkichi Iyanaga, Yukiyosi Kawada. Trans: Kenneth O. May. MIT Pr, 1980. Volume 1, xv + 883 pp; Volume 2, 864 pp, $40 (P) set. [ISBN: 0-262-59010-7] Paperback edition of 1978 English edition (TR, February 1978), at less than one-third the original cost. LAS

General, T??(13: 1). Mathematics, A Liberal Arts Approach with BASIC. Irving Allen Dodes. Krieger, 1980, xiii + 450 pp, $18.50. [ISBN: 0-88275-892-4] Assuming no previous mathematics and intended for the liberal arts college student, this text presents elementary mathematical concepts (including review problems in addition and subtraction) in language suggestive of junior high. This second edition adds a chapter on Basic. JNC

General, S*, P*, L.** The Mathematical Gardner. Ed: David A. Klarner. Wadsworth, 1981, viii + 382 pp. [ISBN: 0-534-98015-5] A stellar cast of Martin Gardner fans (Knuth, Graham, Honsberger, Guy, Berge, Golomb, Schattschneider, Sloane and many others) offer a 65th birthday collage of games (two-pins, hex, mental poker, etc.), geometry (e.g., tangent circles, flexing surfaces), packings and tilings, supernatural numbers, codes, fun and games. A wonderful tribute to mathematics' greatest popularizer. LAS

Precalculus, T(13). Elementary Technical Mathematics, Third Edition. Frank L. Juszli, Charles A. Rodgers. P-H, 1980, xii + 475 pp, $16.95 [ISBN: 0-13-260869-3]; Elementary Technical Mathematics with Calculus, Second Edition, xii + 567 pp, $19.95. [ISBN: 0-13-272732-3] Aimed at technical colleges and industry-operated programs; includes only a bare bones introduction to calculus. (Second Edition, TR, October 1971.) AWR

Precalculus. Analytic Geometry. Charles C. Carico, Irving Drooyan. Wiley, 1980, xiv + 310 pp, $16.95. [ISBN: 0-471-06435-1] Designed to follow a course in college algebra and trigonometry. Covers the topics which traditionally precede calculus. LLK

Precalculus, S(13). Lake Wobegon Math Problems. George Bridgman (Dept. of Math., Univ. of Wisconsin, River Falls, WI 54022), 1979, v + 46 pp, $2 (P). 42 typical precalculus and calculus problems, creatively embedded in the elaborate, imaginative setting of the mythical Lake Wobegon, "the little town time forgot and the decades cannot improve." LAS

Education, P*, L*. An Analysis of Mathematics Education in the Union of Soviet Socialist Republics. Robert B. Davis, et al. ERIC, 1979, iv + 178 pp, $4.25 (P). A series of essays, based both on personal observation and on analysis of published literature, reporting on the nature and content of Soviet mathematics education. A balanced, detailed report that provides needed context for Isaak Wirszup's more recent (and better publicized) report on the same subject. LAS

Education, S(16-17), P. Socialist Mathematics Education. Ed: Frank J. Swetz. Burgandy Pr, 1978, xvi + 421 pp, $12.50 (P); $19.50. [ISBN: 0-917574-04-4; 0-917574-03-6] A survey of mathematics education in socialist countries through seven case studies; Russia, East Germany, China, Hungary, Sweden, Tanzania, Yugoslavia. Includes history of elementary and secondary mathematics education, curricula, teacher training, teaching practices, mathematics learning research and national competitions. Not politically motivated, it seeks to examine influence of social milieu on mathematics education. PJ

Education, P*, L*. Conversations with Jean Piaget. Jean-Claude Bringuier. Trans: Basia Miller Gulati. U of Chicago Pr, 1980, xii + 143 pp, $12.95. [ISBN: 0-226-07503-6] Lightly edited transcripts of 14 interviews conducted between 1969 and 1976 by an award-winning French journalist. Brisk, sophisticated questions and revealing responses provide a vivid portrait of Piaget's conception of his own work. "Everything one teaches a child one prevents him from inventing or discovering." LAS

Education, P. What are the Needs in Precollege Science, Mathematics, and Social Science Education?

Views from the Field. NSF, 1979, xiv + 211 pp, (P). Reports from professional organizations of teachers, of scientists and of administrators on the status of precollege mathematics and science education. Includes a two-part NCTM essay "Mathematics Teaching Today: Perspectives from Three National Surveys" written by Jim Fey. LAS

History, S, P*, L.** John Von Neumann and Norbert Wiener: From Mathematics to the Technologies of Life and Death. Steve J. Heims. MIT Pr, 1980, xviii + 547 pp, $19.95. [ISBN: 0-262-08105-9] A fascinating dual biography of two extraordinary mathematicians, focusing on their professional development, the influence of their ideas on twentieth century science, and their different (but equally vigorous) efforts to influence science policy and politics in the post-Hiroshima era. The similarities and contrasts of their parallel careers, the influence each had on the other, and the psychological roots of their distinctive habits of mind provide unique insight into the nature and accomplishment of genius. LAS

History, P, L. Two Decades of Mathematics in the Netherlands, 1920-1940: A Retrospection on the Occasion of the Bicentennial of the Wiskundig Genootschap. Ed: E.M.J. Bertin, H.J.M. Bos, A.W. Grootendorst. Math Centre, 1978. Part I, xxlv + 193 pp; Part II, 184 pp, Dfl. 42,50 set (P). A series of "close-ups" of Dutch mathematics between the two world wars: 19 excerpts from the writings of influential Dutch mathematicians (e.g., Brouwer, van der Waerden, Heyting, Casimir, van Dantzig, Hurewicz, Freudenthal), each followed by an expert commentary discussing the importance and influence of the passage. LAS

Foundations, P. Lecture Notes in Mathematics-811: Recursion on the Countable Functionals. Dag Normann. Springer-Verlag, 1980, viii + 191 pp, $11.80 (P). [ISBN: 0-387-10019-9] First two chapters introduce functionals of higher types, the subclass of countable (continuous) functionals, Kleene computations, and countable recursions. Later chapters cover topological structure of spaces of countable functionals, degree structure of recursions and computations, envelopes and sections. Assumes knowledge of ordinary recursion theory and elementary descriptive set theory. KS

Foundations, P. Studies in Inductive Logic and Probability, Volume II. Ed: Richard C. Jeffrey. U of Calif Pr, 1980, 305 pp, $20. [ISBN: 0-520-03826-6] Part II of Rudolf Carnap's Basic System of Inductive Logic, picking up where Volume I left off (U. of Calif. Pr., 1971; TR, August-September 1972). Companion essays chosen by the editor present other classic expressions of developments in the field as well as modifications and extensions of Carnap's work. GHM

Foundations, P. Mathematics in the Alternative Set Theory. Petr Vopenka. B.G. Teubner, 1979, 120 pp, 13M (P). "Alternative set theory" (AST) has been developed by Vopenka and others in Prague since 1973. Countering Cantor's theory of actually infinite sets, AST treats infinity "as a phenomenon involved in our observation of large, incomprehensible sets," i.e., potential infinity somewhat in the spirit of the ultra-intuitionist Esenin-Volpin. All sets in AST are classically finite; infinity arises when some sets are postulated to contain subclasses which are not sets. Vopenka rejects the possibility of completely formalizing AST, but notes that certain formalized fragments of it have classical models derived from nonstandard analysis-arithmetic and thus are classically respectable. GHM

Foundations, P. Recursion Theory: Its Generalisations and Applications. Ed: F.R. Drake, S.S. Wainer. London Math. Lect. Note Ser., No. 45. Cambridge U Pr, 1980, 319 pp, $24.50 (P). [ISBN: 0-521-23543-X] Papers based on eleven of the sixteen courses and hour lectures given at the Logic Colloquium held at the University of Leeds, England, August 5-14, 1979. Abstracts of 19 contributed papers appear in Volume 45 of the Journal of Symbolic Logic. JAS

Combinatorics, P. Combinatorics, Representation Theory and Statistical Methods in Groups: Young Day Proceedings. Ed: T.V. Narayana, R.M. Mathsen, J.G. Williams. Lect. Notes in Pure and Appl. Math., V. 57. Dekker, 1980, xii + 170 pp, $25 (P). [ISBN: 0-8247-6937-6] A lengthy article on the elementary theory of the symmetric group, along with several up-to-date research and review articles, all with considerable historical perspective. LCL

Combinatorics, T(16-17), S, P, L. Algorithmic Graph Theory and Perfect Graphs. Martin Charles Golumbic. Comp. Sci. and Appl. Math. Acad Pr, 1980, xx + 284 pp, $29.50. [ISBN: 0-12-289260-7] An introduction to graph theory which emphasizes those problems which are related to the structure of permutation graphs, interval graphs, circle graphs, perfect graphs, and others. Each time a new structure appears the author devotes some effort to a description of algorithms related to that structure. Lots of good exercises. Excellent lists of references. CEC

Combinatorics, P. Combinatorics 79. Ed: M. Deza, I.G. Rosenberg. North-Holland, 1980. Part I, Annals of Discrete Math., No. 8, xxii + 309 pp; Part II, Annals of Discrete Math., No. 9, vii + 309 pp, $65.75 each. [ISBN: 0-444-86112-2] Proceedings of the joint Canada-France combinatorial colloquium held in Montreal, June 11-16, 1979. JAS

Number Theory, S(18), P. Proceedings of the Queen's Number Theory Conference, 1979. Ed: Paulo Ribenboim. Pure and Appl. Math., No. 54. Queen's U, 1980, xiii + 497 pp, $17.50 (P). This collection includes most of the invited lectures and contributed papers presented during the conference. It also includes the series of lectures given by H. Montgomery on the Riemann hypothesis. CEC

Number Theory, T(16-17: 1), S, L. Advanced Number Theory. Harvey Cohn. Dover, 1980, xi + 276 pp, $5 (P). [ISBN: 0-486-64023-X] An unabridged and corrected republication of a 1962 work entitled A

Second Course in Number Theory. Assumes a first course in number theory and elementary calculus. The focus is on ideal theory in quadratic fields and applications of ideal theory. Includes a small number of exercises. CEC

Linear Algebra, T(14: 1). Elementary Linear Algebra. Stewart Venit, Wayne Bishop. Prindle, 1981, viii + 393 pp. [ISBN: 0-87150-300-X] Good organization. Leads the student through linear independence, span, subspaces, basis and dimension in the context of R^n and then introduces abstract vector spaces in Chapter 7. Good selection of applications in Chapters 8 and 9. LLK

Algebra, P. Lecture Notes in Mathematics-778: SK_1 von Schiefkörpern. P. Draxl, M. Kneser. Springer-Verlag, 1980, 123 pp, $10.70 (P). [ISBN: 0-387-09747-3] Proceedings of the 1976 Seminar Bielefeld-Göttingen. JAS

Algebra, P. The Weil Representation, Maslov Index and Theta Series. Gérard Lion, Michèle Vergne. Progress in Math., No. 6. Birkhäuser Boston, 1980, 337 pp, $16 (P). [ISBN: 3-7643-3007-4] Part I treats the Shale-Weil representation of the symplectic group and establishes a relation between its cocycle and the Maslov index. Part II gives applications of theta series to liftings of modular forms. Bibliographical notes, references, no index. JS

Algebra, T(18: 1), S, P. Lecture Notes in Mathematics-789: Arithmetic Groups. James E. Humphreys. Springer-Verlag, 1980, vii + 158 pp, $11.80 (P). [ISBN: 0-387-09972-7] Using Z in R, GL(n,Z) in GL(n,R), SL(n,Z) in SL(n,R), etc., as models, the author defines an arithmetic group as a discrete subgroup of a Lie group defined by arithmetic properties. The aim here is to develop "in an elementary way several of the underlying themes" such as "the construction of a good fundamental domain" and strong approximation. Special attention is given to GL_n and SL_n, the congruence subgroup problem, and related questions. Suggestions for further reading, bibliography, indices. JS

Algebra, S(16-18), P, L. Lecture Notes in Mathematics-790: Groups, Trees and Projective Modules. Warren Dicks. Springer-Verlag, 1980, ix + 127 pp, $9.80 (P). [ISBN: 0-387-09974-3] A self-contained, elementary (no mention of cohomology!) proof of one of the major results in the theory of cohomology of groups: For any nonzero ring R (associative, with 1) and group G, the augmentation ideal of the group ring R[G] is right R[G]-projective if and only if G is the fundamental group of a graph of finite groups having order invertible in R. LCL

Algebra, P. Semigroups. Ed: T.E. Hall, P.R. Jones, G.B. Preston. Acad Pr, 1980, x + 255 pp, $18. [ISBN: 0-12-319450-4] Proceedings of a four day conference on semigroups held at Monash University, Australia, in October 1979. Includes survey articles as well as many previously unpublished results. LCL

Calculus, T(13-14). Calculus with Analytic Geometry. Howard Anton. Wiley, 1980, xxiv + 1221 pp, $26.95. [ISBN: 0-471-03248-4] A fairly standard large three semester (through Green's theorem) calculus book featuring nice computer graphics and much attention in its writing to pedagogical detail. To paraphrase the second paragraph of the author's Preface: Since nobody can cover this much material in three semesters, the book needs to be readable. JAS

Calculus, T(13: 2). Calculus: Single-Variable. Jerrold Marsden, Alan Weinstein. Benjamin/Cummings, 1981, xxvii + 713 pp, $19.95. [ISBN: 0-8053-6936-8] First 13 chapters (through infinite series) of Calculus (TR, May 1980). LAS

Calculus, S*(13), P, L. Calculus Unlimited. Jerrold Marsden, Alan Weinstein. Benjamin/Cummings, 1981, xiii + 235 pp, $6.95 (P). [ISBN: 0-8053-6932-5] A supplement to any standard calculus text, using Eudoxus' method of exhaustion as an alternative to the theory of limits. Rigor is provided by the notion of a "transition point," a concept closely linked to geometry (e.g., inflection points, catastrophe theory), and to certain parts of science (e.g., phase transitions) not normally modelled by traditional calculus. Provides a solid geometric (rather than analytic) basis for the theory of calculus. LAS

Calculus, T(13: 1, 2). Elementary Technical Mathematics with Calculus. James F. Connelly, Robert A. Fratangelo. Macmillan, 1979, xiv + 958 pp, $16.95. [ISBN: 0-02-324440-2] A year course in elementary functions and first term calculus (through techniques and applications of integration), with a practical imitative flavor: each of the 132 sections consists almost entirely of worked examples and exercises. Contains numerous appendices, as well as conversion and formula tables inside both front and back covers. LAS

Complex Analysis, P. Lecture Notes in Mathematics-822: Séminaire Pierre Lelong--Henri Skoda (Analyse) Années 1978/79. Ed: Pierre Lelong, Henri Skoda. Springer-Verlag, 1980, viii + 356 pp, $23 (P). [ISBN: 0-387-10241-8] A wide variety of papers on functions of several complex variables. Some papers from the seminar not included here have appeared elsewhere. JAS

Complex Analysis, P. Deformations of Coherent Analytic Sheaves with Compact Supports. Yum-Tong Siu, Günther Trautmann. Memoirs No. 238. AMS, 1981, iii + 155 pp, $8.80 (P). [ISBN: 0-8218-2238-1] This monograph describes the construction of a semi-universal deformation of any coherent analytic sheaf with compact support. JAS

Differential Equations, P. Nonlinear Partial Differential Equations: Sequential and Weak Solutions. Elemer E. Rosinger. Math. Stud., V. 44. North-Holland, 1980, xix + 317 pp, $39 (P). [ISBN: 0-444-

86055-X] Sequel to author's 1978 Distributions and Nonlinear Partial Differential Equations (TR, February 1980). Author's main concern is "a unified way of dealing with weak solutions of general nonlinear partial differential equations, a way which would exhibit a certain natural, objective trait, thus going beyond the somewhat ad hoc appearance the customary functional analytic methods, with their wealth of linear toplogical structures and function spaces used, might now and then suggest." JK

Differential Equations, T(16-17). Elements of Soliton Theory. G.L. Lamb, Jr. Wiley, 1980, xii + 289 pp, $29.95. [ISBN: 0-471-04559-4] An introductory treatment which presupposes background in eigenvalue problems and some background in complex integration and quantum theory. Differential geometry is not needed. Appears to offer intuition in addition to techniques. JAS

Numerical Analysis, P. Analysis and Computation of Fixed Points. Ed: Stephen M. Robinson. Acad Pr, 1980, ix + 413 pp, $22.50. [ISBN: 0-12-590240-9] Eight of the thirteen papers presented at the symposium held at the University of Wisconsin, Madison, May 7-8, 1979, together with a closely-related doctoral dissertation by Charles Engles. JAS

Numerical Analysis, P. Approximation Theory III. Ed: E.W. Cheney. Acad Pr, 1980, xxiii + 944 pp, $59. [ISBN: 0-12-171050-5] Proceedings of the conference held January 8-12, 1980 at Austin, Texas in celebration of the seventieth birthday of George G. Lorentz. JAS

Functional Analysis, T(18). Linear Operators in Hilbert Spaces. Joachim Weidmann. Trans: Joseph Szücs. Grad. Texts in Math., No. 68. Springer-Verlag, 1980, xiii + 402 pp, $34. [ISBN: 0-387-90427-1] A graduate text which includes applications to mathematical physics. Major topics include closed operators, spectral theory of self-adjoint and normal operators, perturbation theory, operators on $L_2(R^m)$, and scattering theory. Contains about 225 exercises. SG

Functional Analysis, S(18), P. Actions of Finite Groups on the Hyperfinite Type II_1 Factor. Vaughan F.R. Jones. Memoirs No. 237. AMS, 1980, v + 70 pp, $4.40 (P). [ISBN: 0-8218-2237-3] "A complete classification up to conjugacy of the action of a finite group G on the hyperfinite II_1 factor R in terms of three invariants: a normal subgroup N of G; an element of the relative cohomology group $H^2(G/N, G, R/Z)$ and an element of the quotient of a finite dimensional simplex by a simplicial action of $H^1(N, R/Z)^G$ coming from permutations of the vertices." JS

Functional Analysis, T(16-17: 1), S, P, L. Introduction to Functional Analysis: Banach Spaces and Differential Calculus. Leopoldo Nachbin. Trans: Richard M. Aron. Pure and Appl. Math., V. 60. Dekker, 1981, ix + 166 pp, $19.75. [ISBN: 0-8247-6984-8] Aimed toward advanced undergraduates or beginning graduate students. Fairly formal, self-contained, rigorous presentation in two parts: Part I is an introduction to Banach spaces; Part II is a coordinate-free development of differential calculus in normed spaces. Some (generally non-trivial) exercises, bibliography, index. Suitable for self-study. JS

Optimization, T(17: 2), L. A Vector Space Approach to Models and Optimization. C. Nelson Dorny. Krieger, 1980, xix + 599 pp, $34.50. [ISBN: 0-89874-210-2] Uses vector space language and ideas to present a unified treatment of structure and optimization of deterministic models. Many references and comments. JG

Analysis, S(18), P, L. The Radon Transform. Sigurdur Helgason. Progress in Math., No. 5. Birkhäuser Boston, 1980, x + 192 pp, $12 (P). [ISBN: 3-7643-3006-6] In 1917 Johann Radon showed (in a paper reproduced in the appendix to this volume) that a differentiable function on R^3 can be determined explicitly from its integrals over the planes in R^3. These notes outline the theory of such transformations, with applications to partial differential equations, X-ray reconstruction and radio astronomy. LAS

Analysis, P. Twelve Papers in Analysis. Ed: Lev J. Leifman. Amer. Math. Soc. Trans, Ser. 2, V. 115. AMS, 1980, v + 202 pp, $31.60. [ISBN: 0-8218-3065-1] Proceedings of the 1974 Drogobych "winter school" on mathematical programming; most papers are on operator theory. LAS

Differential Geometry, T(14-17), S, L**.** Tensor Geometry, The Geometric Viewpoint and its Uses. C.T.J. Dodson, T. Poston. Pitman, 1979, xiii + 598 pp, $24 (P). [ISBN: 0-273-01040-9] This outstanding multivariable calculus and differential geometry book is now available in paperback at a much more reasonable price. (Hardcover edition, TR, October 1978.) JAS

Geometry, S(13), P. Straight Lines and Curves. N.B. Vasilyev, V.L. Gutenmacher. Trans: Anjan Kundu. MIR Pub, 1980, 197 pp, $4. Contains more than 200 geometric problems dealing with paths of moving points, sets of points satisfying certain conditions and determination of maxima and minima; most are given with solutions and comments. A great source of contest problems. JNC

Geometry, P. The Geometry of the Generalized Gauss Map. David A. Hoffman, Robert Osserman. Memoirs No. 236. AMS, 1980, iii + 105 pp, $6.40 (P). [ISBN: 0-8218-2236-5] "This paper is devoted primarily to the study of properties of the Grassmannian of oriented 2-planes in R^n and to applications of these properties to understanding minimal surfaces in R^n via the generalized Gauss map." JAS

Topology, S(18), P. Z/2-Homotopy Theory. M.C. Crabb. London Math. Soc. Lect. Note Ser., No. 44. Cambridge U Pr, 1980, v + 128 pp, $15.95 (P). [ISBN: 0-521-28051-6] A discussion of symmetry in topology by way of various actions of the cyclic group of order 2; in particular the antipodal

involution of a real vector bundle, doubling, squaring, bilinear forms, and Hermitian K-theory. Considerable background needed. Proofs are not given in detail. Extensive bibliography, index. JS

Probability, S(16-18), P. Lecture Notes in Control and Information Sciences-26: Regenerative Simulation of Response Times in Networks of Queues. D.L. Iglehart, G.S. Shedler. Springer-Verlag, 1980, xii + 204 pp, $15.70 (P). [ISBN: 0-387-09942-5]

Probability, S, P, L. Expert Uncertainty and the Use of Subjective-Probability Models. J.M. Dickey. U of Wales Pr, 1980, 22 pp, 90p (P). [ISBN: 0-7083-0756-6] "There is no alternative calculus competing with the theory of subjective probability for the quantification and regularization of expert opinion." An inaugural lecture stating the case for general education in subjective probability. LAS

Probability, P. Lecture Notes in Mathematics-780: Semi-Martingales sur des Variétés, et Martingales Conformes sur des Variétés Analytiques Complexes. Laurent Schwartz. Springer-Verlag, 1980, xv + 132 pp, $10.70 (P). [ISBN: 0-387-09749-X] Motivated by the fact that a C^2 function of a semimartingale is a semi-martingale and a holomorphic function of a conformal martingale is a conformal martingale, the author undertakes a study of semi-martingales with values in a manifold and conformal martingales with values in a complex analytic manifold. SES

Probability, T(17-18: 1, 2), S, P*. Theory and Applications of Stochastic Differential Equations. Zeev Schuss. Wiley, 1980, xiii + 321 pp, $25.95. [ISBN: 0-471-04394-X] Points of origin, basic theory and wide range of applications of stochastic differential equations that contain white noise. Singular perturbation methods and their role in various areas of science. Phenomena in chemical kinetics, solid-state diffusion, genetics, filtering of signals from noise are modeled. Well-planned and clearly written, beginning with relevant probability theory. Working knowledge of advanced calculus, ordinary and partial differential equations and elementary probability theory are prerequisites. Almost one hundred references, most to journal articles. JK

Statistics, P. Nonparametric Sequential Selection Procedures. H. Büringer, H. Martin, K.-H. Schriever. Birkhäuser Boston, 1980, 488 pp, $24.50. [ISBN: 3-7643-3021-X] New procedures for designing a statistical experiment (selection model) to select alternatives (populations). Divided into two parts: dichotomous-response-(selection)-models and continuous-response-(selection)-models. Good set of references. RSK

Statistics, P. Lecture Notes in Mathematics-821: Statistique non Paramétrique Asymptotique. Ed: J.P. Raoult. Springer-Verlag, 1980, vii + 175 pp, $12.70 (P). [ISBN: 0-387-10239-6] Seven of the twelve lectures presented at the conference held at Rouen, France, June 13-14, 1979. Six of the seven papers concern rank statistics. JAS

Statistics, P. Mathematical Statistics. Ed: Robert Bartoszyński, Jacek Koronacki, Ryszard Zieliński. Banach Center Pub, V. 6. PWN, 1980, 376 pp. [ISBN: 83-01-01493-8] Proceedings of the semester on mathematical statistics held in the fall of 1976 at the Banach Center; 43 papers, all in English, on decision problems, multivariate statistics, analytical methods, and case studies. LAS

Computer Programming, T(13: 1), S. Introduction to TRS-80 Level II BASIC and Computer Programming. Michael P. Zabinski. P-H, 1980, xiv + 186 pp, $14.95; $10.95 (P). [ISBN: 0-13-499970-3; 0-13-499962-2] A well-written, comprehensive introduction to this version of Basic. Includes many good examples and exercises. CEC

Computer Programming, T(16-17). COBOL, A Vehicle for Information Systems. Robert T. Grauer. P-H, 1981, xvi + 432 pp, $18.95. [ISBN: 0-13-139709-5] An integrated treatment of programming and information systems for students in a business curriculum. For the sake of general accessibility, this text is written as a case study of a personnel system. JAS

Computer Programming, P. ORACLS, A Design System for Linear Multivariable Control. Ernest S. Armstrong. Control and Systems Theory, V. 10. Dekker, 1980, ix + 244 pp, $35. [ISBN: 0-8247-1239-0] ORACLS stands for Optimal Regulator Algorithms for the Control of Linear Systems. This book is a revised and updated version of a user's guide originally published by NASA. ORACLS is a Fortran-coded system containing up-to-date numerical linear algebra procedures to implement linear-quadratic-Gaussian methodology. JAS

Computer Science, P. Computer Networks and Their Protocols. D.W. Davies, et al. Wiley, 1979, xv + 487 pp, $53.25. [ISBN: 0-471-99750-1] A treatment which includes graph theoretic optimisation analysis of networks and discussion of security and routing algorithms. JAS

Computer Science, T?(16-17), S, L. Interactive Computer Graphics. B.S. Walker, J.R. Gurd, E.A. Drawneek. Crane Russak, 1975, ix + 160 pp, $19.50. [ISBN: 0-8448-0650-1] This 1980 reprint appears to be unchanged from the original. It tends towards a nuts and bolts approach to graphics languages and electronics. Its age may make it most useful as a readable introductory supplement since the book devotes a relatively large amount of space to the elements which have grown into mind boggling complexity in current graphics devices. JAS

Systems Theory, P. 2-D Systems, An Algebraic Approach. R. Eising. Math. Centre Tracts, No. 125. Math Centrum, 1980, 141 pp, Dfl. 17 (P). [ISBN: 90-6196-198-X] A study, based on the author's dissertation, of rings over a principal ideal domain which aids in the development of algorithms for

manipulating two-dimensional systems, e.g., two-dimensional graphics data for a computer. Readable exposition and a good bibliography make this monograph of some interest as a general introduction to image processing. JAS

Applications (Biology), S(16-18), P*, L. Mathematics of Genetic Diversity. J.F.C. Kingman. CBMS Reg. Conf. in Appl. Math., No. 34. SIAM, 1980, vii + 70 pp, $13 (P). [ISBN: 0-89871-166-5] Given our understanding of Mendelian genetics and Darwinian natural selection, we would expect populations to drift toward a common genetic makeup. Yet, this is not what we observe. What is it in the natural environment that maintains the genetic diversity that we observe? Mathematical models have been very helpful in these studies: one can explore how a population would evolve under various sets of biological assumptions. Typically, this process will clarify and refine the basic assumptions. This monograph explores this topic and develops a new model, based on the recent realization that, at a particular chromosome locus, the number of possible alleles is almost unlimited (the classical theory concentrates on loci with two, three, or four alleles). LCL

Applications (Biology), P. Some Mathematical Questions in Biology. Ed: George F. Oster. Lect. on Math. in Life Sci., V. 13. AMS, 1980, vi + 274 pp, $11.20 (P). [ISBN: 0-8218-1163-0] Proceedings of the 1980 AAAS symposium on mathematical biology: whaling and fishery management, modelling population data, heart valves, morphogenesis, lichen growth. LAS

Applications (Engineering), P. Transactions of the Twenty-Sixth Conference of Army Mathematicians. US Army Research Office (P.O. Box 12211, Research Triangle Park, NC), 1980, xii + 399 pp, (P). Papers presented at the conference held at the U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire, June 10-12, 1980. JAS

Applications (Engineering), P. Annual Review of Fluid Mechanics, V. 13. Ed: Milton van Dyke, J.V. Wehausen, John L. Lumley. Annual Reviews, 1981, 530 pp, $20 (P). [ISBN: 0-8243-0713-5]

Applications (Physics), P, L.** Some Strangeness in the Proportion: A Centennial Symposium to Celebrate the Achievements of Albert Einstein. Ed: Harry Woolf. A-W, 1980, xxxi + 539 pp, $43.50. [ISBN: 0-201-09924-1] Complete proceedings (presentations, lectures, discussion, photographs) of a March 1979 symposium at the Institute for Advanced Study, featuring such speakers as Stephen Hawkins, S.S. Chern, Tullio Regge, John Wheeler, and others, on relativity, quantum theory, cosmology, quantum gravity, and the unity of physics. An impressive overview of contemporary theoretical physics. LAS

Applications (Physics), P. Advances in Twistor Theory. Ed: L.P. Hughston, R.S. Ward. Pitman, 1979, 335 pp, $15 (P). [ISBN: 0-8224-8448-X] A collection of fifty short articles adapted from the Twistor Newsletter written by eighteen authors over a three year time span which are loosely gathered into a five chapter introduction to twistor theory. Although the audience is intended to include more than the devotees of the Twistor Newsletter (the volume includes introductory essays), the presentation assumes a thorough knowledge of differential geometry and physics. JAS

Applications (Physics), P*, L.** Encyclopedia of Physics. Ed: Rita G. Lerner, George L. Trigg. A-W, 1981, xvi + 1157 pp, $99.50. [ISBN: 0-201-04313-0] Approximately 500 concise articles spanning all of physics, from Absorption Coefficient to the Zeeman Effect. An extraordinary volume, intelligible to the non-specialist, yet of real value as a reference to any mathematician or scientist. While most articles minimize the use of mathematics, the volume includes separate articles on catastrophe theory, eigenfunctions, ergodic theory, field theory, Fourier transforms, group theory, Kepler's laws, Lie groups, matrices, Newton's laws, operators, statistics, SU(3), vector and tensor analysis. LAS

Applications (Social Science), S, P, L. Arrow's Theorem: The Paradox of Social Choice. A Case Study in the Philosophy of Economics. Alfred F. MacKay. Yale U Pr, 1980, ix + 143 pp, $14.50. [ISBN: 0-300-02450-9] A philosophical attempt to explain the paradox of Arrow's "Impossibility" Theorem on social choice. The author first rejects classical responses (e.g., restricting the scope of majority voting) as unmotivated and ad hoc. He concludes, however, by reformulating Arrow's paradox as an infinite regress, and then rationalizing the classical restrictions as special instances of the well-established tradition of positing a first cause to eliminate the paradox of infinite regress.

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

University of Texas, Austin: *George Lorentz* became Emeritus Professor on September 1, 1980. *Michael Starbird* was promoted to Associate Professor. *Ronald M. Dotzel* has accepted a position with the University of Missouri at St. Louis. *H. Elton Lacey* has resigned to become chairman of the Mathematics Department at Texas A&M University. *Dan Velleman* from the University of Wisconsin has been appointed Instructor. *Darwin E. Peek* is the State of Texas visiting Faculty Honoree for 1980-81. Emeritus Professor Hyman J. Ettlinger, one of fourteen surviving charter members of MAA, was well enough on September 1, 1980, to celebrate his 91st birthday and 62nd wedding anniversary by attending a football game.

South Dakota State University: Professor *Kenneth L. Yocom* has been named Head of the Department of Mathematics. Associate Professor *Larry F. Bennett* has been promoted to Professor and Assistant Professor *Daniel C. Kemp* has been promoted to Associate Professor. Professor *Gerald E. Bergum* has been named Editor of the Fibonacci Quarterly.

Ohio University: *Donald O. Norris*, Chairman of the Department of Mathematics, has been appointed a University Professor. Associate Professor *Hari Shankar* has been promoted to Professor. *William E. Kaufman* of the University of Houston was named Assistant Professor.

Mohawk Valley Community College: Professor *Richard A. Meili* has been appointed Head of the Department of Mathematics. Associate Professor *Ivan Doszpoly* is in Japan for a second year to research the history of mathematics in Japan.

California State University, Fullerton: Recent appointments to the faculty include *Steven Roman* and *Lawrence R. Weill*.

Bronx Community College: Associate Professor *Norman Gore* has been promoted to Professor. *Susan Forman* and *Madeleine Bates* have been promoted from Lecturer to Assistant Professor.

University of Lowell: *Kenneth M. Levasseur* has been appointed Assistant Professor. *Michael Grossman* was promoted to Associate Professor.

Bates College: Associate Professor *Richard W. Sampson* was promoted to Professor. Associate Professor *David C. Haines* has been named Chairman of the Department of Mathematics.

Richard A. Zalik, Assistant Professor at Auburn University has been promoted to Associate Professor.

Carol Ulsafer has left North Central College to accept a position at the University of Montana.

Assistant Professor Sr. *Mary O'Malley* has been promoted to Associate Professor at Rosary College.

Assistant Professor *Ellen Cunningham*, S.P., has been promoted to Associate Professor at Saint Mary-of-the-Woods College.

Professor *William E. Dorgan* of Western State College at Gunnison, Colorado, has retired with emeritus rank.

Kimball Hughes, formerly at UCLA, has been named Visiting Assistant Professor at Reed College.

Professor *Ronald Butler* has retired from Edinboro State College and has been appointed Professor at Minot State College.

Assistant Professor *Phillip E. Parker* has left Syracuse University to accept a position at the University of Missouri.

Associate Professor *William Fleming Stout* of the University of Illinois has been promoted to Professor.

Irl Bivens, instructor at Pfeiffer College, has left to do post-doctoral work at Rice University

Michael Jacobs, formerly chairman of the Department of Mathematics at Cypress College, is now Chairperson of its Division of Science, Engineering and Mathematics.

Professor *O.D. Harrold* has retired from the faculty of Florida State University.

Assistant Professor *Thomas M. Thompson* of Walla Walla College has been promoted to Associate Professor.

Kansas State University: *Floyd B. Sloat* recently retired with the rank of Emeritus Associate Professor. Assistant Professor *Thomas B. Muenzenberger* has been promoted to Associate Professor.

University of Tennessee: Professor *Lida K. Barrett*, Head of the Department of Mathematics, has become Associate Provost of Northern Illinois University. Associate Professor *David E. Dobbs* has been promoted to Professor.

University of New Mexico: Associate Professors *Gustave Efroymson* and *Cornelius Onneweer* have been promoted to Professor. Associate Professor *James V. Lewis* has retired.

Lake Superior State College: Assistant Professors *Thomas Mickewich* and *Bernard Arbic* have been promoted to Associate Professor.

East Texas State University: *Allan Avery* has resigned to take a position with Lincoln Land Community College, Associate Professor *John F. Lamb* has been promoted to Professor.

Emory University: Assistant Professor *Beth A. Barron* has resigned to take a similar position at Oakland University. Professor *Trevor Evans*, Chairman of the Department, has been appointed to the Fuller Chair of Mathematics.

Georgetown University: Assistant Professor *James T. Sandefur* was promoted to Associate Professor. Assistant Professor *Ronald C. Rosier* was named Acting Chairman of the Department of Mathematics.

University of Hawaii at Manoa: Assistant Professors *L. Thomas Ramsey* and *Leslie C. Wilson* have been promoted to Associate Professor.

Livingston University: Associate Professor *David Cochener* has resigned to accept a position at Angelo State University. Assistant Professor *G.M. Reokie* was promoted to Associate Professor.

Pierre J. Malraison, Jr., has been promoted to Chief Staff Mathematician at Manufacturing Data Systems, Inc., Ann Arbor.

Assistant Professor *Jack Diamond* has been promoted to Associate Professor at Queens College, CUNY.

Associate Professor *Francis G. Florey*, Chairman of the Department of Mathematical Sciences at the University of Wisconsin-Superior, has been promoted to Professor.

Professor *I.H. Rose* has been elected Chair of the Mathematics Department at Lehman College, CUNY.

Assistant Professor *Jacqueline Dewar* has been promoted to Associate Professor at Loyola Marymount University.

Paul R. Patten, formerly at Brewton-Parker College, has been named Assistant Professor at North Georgia College.

Professor *Theodore G. Ostrom* of Washington State University was chosen to give the Distinguished Faculty Address in November, 1980. The honor was accompanied by an honorarium of $1000.

Andreas N. Phillipou, formerly Associate Professor at the American University of Beirut, is now Professor of the Chair B of Applied Mathematics at the University of Patras-Greece.

Judith M.S. Prewitt, of the Division of Computer Research and Technology, National Institutes of Health, has been elected Fellow of the Institute of Electrical and Electronics Engineers (IEEE).

Concordia College: *Janice J. Vandever*, formerly Assistant Professor at Moorhead State University, has been named Assistant Professor. Assistant Professor *James L. Forde* has been promoted to Associate Professor.

Colorado College: Associate Professor *Daniel J. Sterling* has resigned to take a position with the Martin Marietta Corporation. Instructor *John J. Watkins* has been promoted to Assistant Professor.

Mort Goldberg has been appointed Chairman of the Department of Mathematics at Broome Community College.

COMING SOON IN THE MONTHLY

The following articles will appear in the AMERICAN MATHEMATICAL MONTHLY for May 1981:

Philip J. Davis, *Are There Coincidences in Mathematics?*
W. Hrusa and J.L. Troutman, *Elementary Characterization of Classical Minima*
H.B. Griffiths, *Cayley's Version of the Resultant of Two Polynomials*
Solomon W. Golomb, Corrections to *"Cyclotomic Polynomials and Factorization Theorems"*

The following articles are among those which the editors have accepted for later issues of the MONTHLY. The order of listing does not indicate the order in which they will appear.

William Abikoff, *The Uniformization Theorem*
W. Brian Arthur, *Why a Population Converges to Stability*
John Bloom and Lee Whitt, *The Geometry of Rolling Curves*
R.P. Boas, *Can We Make Mathematics Intelligible?*
R.B. Burckel, *Iterating Analytic Self-maps of Discs*
Roger L. Cooke, *Almost Periodic Functions*
L.E. Dubins and D.A. Freedman, *Machiavelli and the Gale-Shapley Algorithm*
Solomon W. Golomb, *Irrational Sums and Twin Primes*
I. Grattan-Guinness, *On the Development of Logics Between the Two World Wars*
Frank D. Grosshans, *Rigid Motions of Conics*
John Holbrook, *Stochastic Independence and Space-filling Curves*
Patricia Kenschaft, *Black Women in Mathematics in the United States*
Edith H. Luchins, *Women and Mathematics: Fact and Fiction*
Saunders MacLane, *Mathematical Models--A Sketch for the Philosophy of Mathematics*
Anthony Ralston, *Computer Science, Mathematics and the Undergraduate Curricula in Both*
Karen D. Rappenport, *S. Kovalesky*
Michael Rosen, *Abel's Theorem on the Lemniscate*
Nura D. Turner, *Twenty Years Later: High School Students Who Showed Promise in Mathematics*
Edward T. Wong, *Polygons, Circulant Matrices, and Moore-Penrose Inverses*

MAY MEETINGS IN MADISON

The University of Wisconsin-Madison will host a Topical Review of Computation and Analysis of Reacting Flows May 11 and 12, 1981, and a symposium on Transonic Shock and Multidimensional Flows May 13-15.

The symposium will feature lectures by fifteen experts in these fields of current technological interest. The preceding topical review will describe work, mainly by army scientists, on combustion, detonation, and related subjects.

For further information on the symposium contact Mrs. Gladys Moran, Mathematics Research Center, 610 Walnut Street, Madison, WI 53706. For more information on the Topical Review, contact Dr. J. Chandra, Mathematics Division, U.S. Army Research Office, P.O. Box 12211, Research Triangle Park, NC 27709.

1981 CONFERENCE OF NYSMATYC

The New York State Mathematics Association of Two Year Colleges will hold its annual conference at John's Niagara Hotel in Niagara Falls, New York, April 22-26, 1981. All interested persons are invited to attend. Further information is available from Bruce F. Haney, Mathematics Department, Onondaga Community College, Syracuse, NY 13215.

TIME SERIES MEETING

The Fifth International Time Series Meeting is being held in Houston, Texas, August 6-7, 1981. Papers, program participants, and program suggestions are solicited. Papers are sought in both theoretical and applied areas, including the biological, environmental, engineering, economic, and social sciences. A proceedings volume will be published. Persons desiring to present a paper should submit a one-page abstract by May 1, 1981. Send all enquiries and submissions to M. Ray Perryman, Director, Center for the Advancement of Economic Analysis, Baylor University, Waco, TX 76798.

NUMERICAL ANALYSIS CONFERENCE IN DUBLIN

NASECODE II, the second international conference on the Numerical Analysis of Semiconductor Devices and Integrated Circuits, will be held June 17 to 19, 1981, in Dublin, Ireland. The sponsors are IEEE (Electron Devices Society), IEE (Irish Branch), Royal Irish Academy and Irish Mathematical Society. All communications should be addressed to NASECODE II, 39 Trinity College, Dublin 2, Ireland.

DETERMINISTIC AND STOCHASTIC SCHEDULING INSTITUTE

An Advanced Study and Research Institute of Theoretical Approaches to Scheduling Problems will be held in Durham, England, from July 6 to July 17, 1981. The Institute is sponsored by the NATO Advanced Study Institutes Programme and Systems Science Panel, by the Institute of Mathematics and Its Applications, and by the Mathematisch Centrum, Amsterdam. Lecturers include M.A.H. Dempster, E. Gelenbe, E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan (Program Committee), E.G. Coffman, Jr., M.L. Fisher, J.C. Gittins, S.M. Ross, L.E. Schrage, and G. Weiss. Further information can be obtained from J.K. Lenstra and A.H.G. Rinnooy Kan, C/o Econometric Institute, Erasmus University, P.O. Box 1738, 300 DR Rotterdam, The Netherlands.

THE SMITHSONIAN SCIENCE INFORMATION EXCHANGE, INC.

The Smithsonian Science Information Exchange, a non-profit corporation of the Smithsonian Institution, has announced the availability of information packages on research progress in geometry and topology. There are five different packages, consisting of sets of one-page descriptions of on-going and recently completed projects. For prices and other information, write to the exchange at Room 300, 1730 M Street, N.W., Washington, D.C. 20036.

PUBLICATIONS ON COMPUTERS IN EDUCATION

The International Council for Computers in Education (ICEE) issues several publications of value to elementary or secondary school teachers and to teachers of teachers. These include the Computing Teacher (7 issues a year) and booklets in the Instructional Use of Computers and on Computers in the elementary school. For information write to David Moursund, Computer and Information Sciences Department, University of Oregon, Eugene, OR 97403.

NEW PROGRAM AT THE UNIVERSITY OF OKLAHOMA

The University of Oklahoma has established a Master of Business Administration--M.S. in Mathematics dual degree program. Graduates receive both degrees, but up to nine credit hours of course work in each degree program can count toward the other. The program is designed to meet the increasing need for managers and executives with a sophisticated knowledge of modern and classical mathematics.

For more information contact the Department of Mathematics, University of Oklahoma, 601 Elm Avenue, Norman, OK 73019

SUGGESTION BOX

Members of the MAA are encouraged to send in suggestions, questions, etc., about the operation of the Association. Communications will be referred to the appropriate officer of the Association for answering; from time to time those of general interest may also be answered in one or both of the official journals. Communications should be addressed to: Suggestion Box, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036.

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

ANNUAL MEETING OF THE EASTERN PENNSYLVANIA AND DELAWARE SECTION

The Eastern Pennsylvania and Delaware Section of MAA met at the University of Delaware on November 22, 1980. Officers elected are: President, *Howard Anton* (1981); Vice-President, *Bing Wong* (1981) and Members-at-large of the Executive Committee: *Marialusia McAllister* (1983) and *Robert Murphey* (1983).

Invited talks were given by Prof. *Mark Kac*, "Recollections and Reflections on Fifty Years of Probability Theory;" Dr. *Ward Whitt*, "Approximation for Networks of Queues (Description of Complex Systems Adequate for Engineering Purposes);" Prof. *Richard Anderson*, "Algorithmically Defined Functions."

FALL MEETING OF THE MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA SECTION

The Maryland-District of Columbia-Virginia Section of the MAA met November 14 and 15, 1980, at Goucher College in Towson, Maryland. Approximately 140 persons attended this first two-day session for the section.

The invited address was given by Dr. *Richard MacCamy* of Carnegie-Mellon University. His topic was "The Applied Mathematics Program at Carnegie-Mellon University: A Blend of Traditional and Professional Training." Following the banquet and address were two "Birds-of-a-Feather" sessions: "What's Happening at the Two Year Colleges?" led by *William Sweeter*, Montgomery College, and "Preparing Students for a Career as a Nonacademic Mathematical" led by *Marjorie Stein*, U.S. Postal Service.

Three program sessions ran concurrently Friday afternoon and Saturday morning. The following were contributed:

Opportunities in Mathematics for Women, Minorities, and the Handicapped, Martha Ross Redden, American Association for the Advancement of Science
Election Year Mathematics, Robert Lewand, Goucher College
Numerical Analyses: Practice and Pitfalls, Diane O'Leary, University of Maryland
Approximation of Functions of Several Variables, Caren Diefenderfer, Hollins College
Comparison of Some Mathematical Programming Techniques: Linear Programming Software Testing, Ernest Maybrey, Department of Energy
Compacted Equipartition Yields, Planck's Law, Etc., H. John Hays, Naval Research Laboratory
Maximum Likelihood Estimation of Parameters Using Data Grouped in Overlapping Intervals, J. Van Bowen, University of Richmond
Implicit Enumeration of the Resource Control Problem Using T-Sets, Wilfred Candler, World Bank
A Path-Following Method for Obtaining a Connected Component of the Solution Surface of a Connection-Diffusion Model in Fluid Flow, Raymond Mejia, National Institutes of Health
Application of Ridge Regression in Fishery Statistics, Ronald Dick and Sudhamay Basu, Chi Associates Inc.
A Real Quarter-World Application of Mathematics, Coline Makepeace, Federal Emergency Management
A Linear Programming Model for Optimal Protocol Design, J.F. Heafner and Fran Nielson, National Bureau of Standards
A New System for Traffic Simulation, Guido Radelat and George Tiller, Federal Highway Administration
How to Win at Chuck-a-Luck, Michael Chamberlain, U.S. Naval Academy
Mathematical Foundations of Computer-Assisted Analysis, William Kurator, Department of Energy
The New Russian Linear Programming Algorithm, James Falk, George Washington University
Calculus with the Computer, Ted Benac, U.S. Naval Academy
Developing Classroom Applications of Mathematics at the University of Maryland, Mary Myerson, University of Maryland.
The Interdisciplinary Program in Applied Mathematics at the University of Maryland, Peter Wolfe, University of Maryland
Introductory Probability and Statistics with the Computer, John Turner, U.S. Naval Academy
Conflict Monitoring Model for Route Spacing in the High Altitude Airspace, Arthur Smith, the MITRE Corporation
Sensitivity Analysis of the High Altitude Conflict Monitoring Model, Dana Hall, the MITRE Corporation
Determining, Efficient Transportation Patterns of Strategic Materials During Industrial Mobilization, Michael Pfetsch, Federal Emergency Management Agency
An Algorithm to Route Jets for Transporting Checks Among Federal Reserve Banks, David McCarthy, Department of Labor, and Y.C. Park, Naval Sealift Command
A Model to Determine Operational Efficiency of Transit Vehicles on Simulated Bus Routes, William Magro and J.L. Huttingar, Booz, Allen, and Hamilton, Inc.
A Simplified Game Theory Model for the Optimum Alarm Level from Nuclear Material Accounting Data, Martin Messinger and S. Mcgleever, U.S. Nuclear Regulatory Commission.

The business meeting concluded the activities on Saturday. Section Chairman John Smith reported on two section workshops run last summer and announced two for summer 1981 at Salisbury State College Salisbury, Maryland: H.T. Odum and E.C. Odum's *Energy Systems Modeling,* 1-5 June, and Alan Tucker's *Combinatorial Problem Solving,* 8-12 June. Each workshop will cost $150. A report was also given on the Annual High School Mathematics Examination indicating that 17,988 examinations were given in the MD-DC-VA section and that two $500 scholarships, one to Sweet Briar College and one to V.P.I. & S.U., had been added to the awards list.

Robert Hanson, Secretary

CALENDAR OF FUTURE MEETINGS

Sixty-first Summer Meeting, Pittsburgh, Pennsylvania, August 17–19, 1981.
Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Duquesne University, Pittsburgh, Pennsylvania, May 15–16, 1981.

EASTERN PENNSYLVANIA AND DELAWARE, Pennsylvania State University, Ogontz Campus, April 4, 1981.

FLORIDA, early March. Deadline for paper titles two weeks before meeting.

ILLINOIS, Illinois State University, Normal, May 1–2, 1981.

INDIANA, Indiana University–Purdue University, Indianapolis, April 11, 1981.

INTERMOUNTAIN, Brigham Young University, Provo, Utah, April 10–11, 1981.

IOWA, Coe College, Cedar Rapids, April 24–25, 1981.

KANSAS, Benedictine College, Atchison, April 10–11, 1981.

KENTUCKY, Jefferson Community College, Louisville, April 3–4, 1981.

LOUISIANA-MISSISSIPPI, Friday-Saturday before February 20. Deadline for papers three months before meeting.

MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA, William and Mary College, Williamsburg, Virginia, April 11, 1981.

METROPOLITAN NEW YORK, Lehman College, CUNY, May 2, 1981.

MICHIGAN, Oakland University, Rochester, May 1–2, 1981.

MISSOURI, Northwest Missouri State University, Maryville, April 10–11, 1981.

NEBRASKA, University of South Dakota, Vermillion, South Dakota, April 10–11, 1981.

NEW JERSEY, early November, and early May.

NORTH CENTRAL, Mankato State University, Mankato, Minnesota, May 1–2, 1981.

NORTHEASTERN, New England College, Henniker, New Hampshire, June 12–13, 1981.

NORTHERN CALIFORNIA, first or second Saturday in February.

OHIO, Miami University, Oxford, April 10–11, 1981.

OKLAHOMA–ARKANSAS, (approx.) Friday and Saturday of first weekend in April. Deadline for papers three weeks before meeting.

PACIFIC NORTHWEST, Lewis and Clark College, Portland, Oregon, June 19–20, 1981.

ROCKY MOUNTAIN, Colorado College, Colorado Springs, May 1–2, 1981.

SEAWAY, Syracuse University, Syracuse, New York, April 10–11, 1981.

SOUTHEASTERN, University of Alabama, Birmingham, April 10–11, 1981.

SOUTHERN CALIFORNIA, first or second Saturday in March.

SOUTHWESTERN, New Mexico State University, Las Cruces, April 3–4, 1981.

TEXAS, San Antonio College, San Antonio, April 10–11, 1981.

WISCONSIN, Friday and Saturday between mid-April and first week in May. Deadline for papers six weeks before meeting.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.

AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES

AMERICAN MATHEMATICAL SOCIETY, Pittsburgh, Pennsylvania, August 18–21, 1981.

AMERICAN SOCIETY FOR ENGINEERING EDUCATION

ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.

ASSOCIATION FOR SYMBOLIC LOGIC

ASSOCIATION FOR WOMEN IN MATHEMATICS

CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES

FIBONACCI ASSOCIATION

INSTITUTE OF MATHEMATICAL STATISTICS

MU ALPHA THETA

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, St. Louis, Missouri, April 22–25, 1981.

OPERATIONS RESEARCH SOCIETY OF AMERICA, Four Seasons Sheraton, Toronto, Canada, May 4–6, 1981.

PI MU EPSILON

SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.

CONTENTS

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

MAY

THE AMERICAN MATHEMATICAL MONTHLY

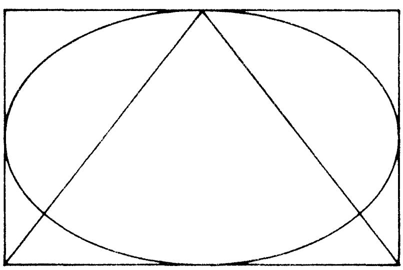
THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 88, Number 5

Coincidences in Mathematics

Classical Minima

Resultant of Polynomials



Honoring Archimedes
(see p. 339)

Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 5**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

ARE THERE COINCIDENCES IN MATHEMATICS?

PHILIP J. DAVIS
Division of Applied Mathematics, Brown University, Providence, RI 02912

This, too, is probable, according to that saying of Agathon: "It is part of probability that many improbable things will happen."
—Aristotle: *Poetics*

The twentieth century may be partially characterized as far as mathematics is concerned as the century of intensive abstraction. The genesis and the program of abstraction was summed up in the first decade of the century by E. H. Moore of the University of Chicago, who wrote:

> The existence of analogies between the central features of various theories implies the existence of a general theory which underlies the particular theories and unifies them with respect to their central features.

Putting it slightly differently, abstraction perceives the common features of diverse theories and fits them into a superstructure wherein they will be unified with respect to this common feature. There was clear apprehension on Moore's part as to what was happening and a prophetic vision and zeal as to what would happen.

As common examples of abstraction, one may cite the algebraic structures of all sorts: semigroups, groups, rings, fields, manifolds, and abstract spaces, such as Hilbert and Banach spaces. But there are many others. As I write, I have before me books on abstract optimality theory, abstract harmonic analysis, and so on. The whole program of abstraction has been so pervasive and so influential that it would not miss the mark by much to say that, in the view of many contemporary mathematicians, mathematics is precisely the study of abstract structures of all sorts.

Returning to Moore's observation that diverse theories are unified with respect to their *common features*, we may now raise the embarrassing question: What is a common feature of two diverse theories? How does one recognize such a feature? And having recognized such a feature, from whence does one derive the confidence that it is a worthwhile task to unify along such a feature? This is not a question of mathematics as such, but a question in the history, psychology, aesthetics, and applications of mathematics, and ultimately it is answerable only in terms of a mathematical culture and certain values that are operative in it. I know of no way of defining—before the act—what a "common feature" means; this is allied to the old philosophical question of universals. But the act of abstraction does it for me. This is precisely one of its functions.

Having raised the question of possible definitions, I should like to twist Moore's quotation still further and ask: Is it possible that a common feature is a coincidence of form which one has observed in several different places? And I mean here to stress the word *coincidence* almost in its surprise aspect. I think that one variety, at least, of a common feature may be identified with coincidence.

I cannot define coincidence. But I shall argue that coincidence can always be elevated or organized into a superstructure which performs a unification along the coincidental elements. The existence of a coincidence is strong evidence for the existence of a covering theory.

Philip J. Davis received his Ph.D. from Harvard under Ralph Boas. He has taught at Harvard, Maryland, the University of Utah, and Brown. He has extensive industrial and government experience, including five years as Chief, Numerical Analysis Section, National Bureau of Standards; he was a Guggenheim Fellow in 1956–57. His extensive work in numerical analysis and applied mathematics includes the books *Interpolation and Approximation* (1963), *Mathematics of Matrices* (1964), *Numerical Integration* (with P. Rabinowitz, 1967), *Circulant Matrices* (1979). A new book, *The Mathematical Experience*, written jointly with Professor Reuben Hersh of the University of New Mexico, appeared recently. Professor Davis received the 1960 Award in Mathematics of the Washington Academy of Sciences and the MAA Chauvenet Prize in 1963.—*Editors*

I shall elucidate the experience of coincidence with a number of examples. The examples selected are formed, for the most part, from the simplest of mathematical ingredients.

Example 1. First, a personal experience. In the eighth grade of elementary school I learned that the diagonal of a rectangle was related to the sides by the Formula of Pythagoras,

$$d^2 = a^2 + b^2.$$

As an adjunct, I learned to carry out the laborious square root algorithm with pencil and paper. A week or so later, I learned that the diagonal of a three-dimensional box was related to the sides of the box by the formula

$$d^2 = a^2 + b^2 + c^2.$$

I recall distinctly that the feeling came over me that this was a great coincidence. Why was the formula not something like

$$d^2 = a^2 + b^2 + c^2 + (abc)^{2/3}?$$

There were plenty of difficult forms knocking around in my textbook. (The formula just given reduces to the two-dimensional case when a, b, or $c = 0$ and is dimensionally correct. But of course it does not reduce properly on all subspaces, which is precisely how the correct formula is obtained.) Moreover, I was grateful that on account of this coincidence I would not have to learn an additional laborious algorithm in order to compute the diagonals of a box. Not so if one wants the side of a cubic box, given its volume.

From the point of view of a mature mathematician, the use of the word *coincidence* in this connection might be regarded as demeaning. What we have here, he might say, is no *mere coincidence*. It is a general theory that extends all the way from the rectangle to the box, from the box to the hyper box, to the box in countably infinite dimensional Hilbert space, to the box in nonseparable spaces, to the ... It is part of a great, useful, expansible theory, part of a Grand Plan, part of the Decreed Order of the Universe. He might assert all this, forgetting for the moment that he has had to temper the Hilbert universe in such a way that the box makes sense:

$$a^2 + b^2 + c^2 + \cdots < \infty,$$

and that some authorities advocate regarding the theorem as a *definition*. (See [**9**, p. 36].)

In this view, to talk about coincidences, *mere* coincidences, one would have to fasten upon a parallelism that is totally "accidental," "uncaused," something paltry, something almost to be disregarded and certainly not worthy of being honored with a proof.

Example 2. The numbers π and e are the outgrowths of substantial, but different, theories. Write down the decimal expansion of π and e:

Number of Digits

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
π	3	1	4	1	5	9	2	6	5	3	5	8	9	7	9	3
e	2	7	1	8	2	8	1	8	2	8	4	5	9	0	4	5

Notice the coincidence in the 13th digit of π and the 13th digit of e (and no earlier digit). I call this a mere coincidence, because our first reaction upon seeing it may very well be "So what!" This, despite the number-mystical fact that the coincidence occurs at the 6th prime. I can call it a coincidence despite the fact that the statement

"The thirteenth decimal digits of both π and e are identical"

is, in fact, a mathematical theorem and can be proved rigorously by established methods of mathematical proof.

As we all know, it is possible in different ways to elevate this *mere* coincidence into something more substantial. Let us look for further coincidences:

Number of Digits

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
π	2	3	8	4	6	2	6	4	3	3	8	3	2	7	9	5	0	2	8	8
e	2	3	5	3	6	0	2	8	7	4	7	1	3	5	2	6	6	2	4	9

In this way, we have discovered further coincidences at the 17th, 18th, 21st, and 34th digits. Our experience with this kind of problem now tells us that it is extremely unlikely that we can discover an elementary formula for the location of the nth coincidence; but exposure to probabilistic thinking, which is designed specifically to elevate chaos and mere coincidence to a Grand Plan, would lead us to say that we observed 5 coincidences in 36 observations. If, as seems likely, the digits of π and e are distributed at random and wholly independently from among $0,1,2,\ldots,9$ (again, which seems likely psychologically, if not logically), then we would expect, on average, one coincidence every ten digits. Since the digits of π and of e are now available to hundreds of thousands of decimal places, a devotee of this problem might like to count the coincidences and subject them to a statistical analysis via the binomial distribution. (In a later example we refer to a similar investigation.) At any rate, the statement

"The decimal digits of π and e coincide, on average, once every ten digits"

is, to my knowledge, neither proved nor disproved. It is difficult of proof or disproof, but is credible and psychologically satisfying. Furthermore, the statement and the denial of the statement are both interesting; and if the denial turned out to be true, we would be inclined to decorate it with an exclamation point.

Example 3. Consider the formula

$$x=\sqrt{1141y^2+1}\,.$$

For $y=1,2,3,\ldots,100$, it is an easy job to show on a hand-held computer that the corresponding x is *not* an integer. We may be led to wonder whether x is ever an integer. The truth of the matter is that it is not an integer for $1\leqslant y\leqslant 10^{25}$. The first value of y for which x is an integer is

$$y=30{,}693{,}385{,}322{,}765{,}657{,}197{,}397{,}208.$$

Now we have no *direct* computational experience of the dearth of integer solutions in the range $1\leqslant y\leqslant 10^{25}$. The anomaly is explained within the Theory of the Pell Equation, which informs us that if the (periodic) continued fraction expansion of $\sqrt{d}$ has a long period then the first solution of the Diophantine equation

$$x^2-dy^2=1$$

will be exceedingly large. The solution just quoted has been arrived at backwards via a comprehensive theory whose existence could have been surmised from experiments in the more modest range $1\leqslant y\leqslant 1000$.

(See [**19**].)

Example 4. Consider the two numbers

$$A=\sqrt{5}+\sqrt{22+2\sqrt{5}}$$

$$B=\sqrt{11+2\sqrt{29}}+\sqrt{16-2\sqrt{29}+2\sqrt{55-10\sqrt{29}}}\,.$$

I ask you now to go to a computer and, using a language that does multiple precision arithmetic,

compute the numbers A and B. If you do, you will find that A and B agree at the very least to 26 significant figures:

$$7.38117\ 59408\ 95657\ 97098\ 72669.$$

Now either (1) This is an incredible coincidence, or (2) $A = B$ is an incredible identity, inasmuch as A and B do not appear "to the naked eye" to lie in the same algebraic field.

If $A \neq B$, then we would still be left with the necessity of elevating the coincidence by placing it within a more comprehensive theory. After all, you just cannot throw integers and square roots around at random and expect to get 26-figure coincidence. There would have to be something behind the coincidence; and, whatever it is, that something would probably enable us to construct further incredibilities of a similar sort.

The reader may infer from the title of one of the references whether $A = B$ or $A \neq B$.

I first heard of this and similar coincidences through H. O. Pollak of the Bell Telephone Laboratories. His interest in the matter lay in an auxiliary question. Let A and B be irrational algebraic numbers. How can one prove or disprove the equation $A = B$ by systematic rational operations? (For example, the identity

$$1 + \sqrt{3} = \sqrt{3 + \sqrt{13 + 4\sqrt{3}}}$$

can be established by two squarings) and can this be carried out in polynomial time? At the time of writing, the last question had not yet been resolved.

(See [**15**], [**18**].)

Example 5. In elementary plane geometry, there are numerous theorems which contain a strong flavor of coincidence. Thus:

(a) the three medians in any triangle intersect in a common point.

Not only do we have this but

(b) the three angle bisectors intersect in a point, and
(c) the three altitudes intersect in a point.

Each of these circumstances represents a coincidence. Many observers view (a), (b), (c) as a further coincidence since their conclusions are identical: "They intersect in a point." This experience may be part of the reason for elevating the separate instances into a comprehensive theory:

CEVA'S THEOREM. *Let the sides of a triangle ABC be divided at L, M, N in the respective ratios* $\lambda : 1$, $\mu : 1$, $\nu : 1$. *Then the three lines AL, BM, CN are concurrent if and only if* $\lambda\mu\nu = 1$.

Now the ideas of (a), (b), (c) are easily extended to higher dimensional simplexes. What about the truth of the resulting statements? Focus attention on (c). Do the altitudes of a tetrahedron intersect? This is a problem about which the Harvard geometer J. L. Coolidge once wrote that most highly educated mathematicians do not know the answer. On the one hand, direct spatial intuition fails most of us. We simply do not have enough experience playing with wire tetrahedra and dropping perpendiculars. On the other hand, if one argues by verbal analogies, one is inclined to answer the question with "yes." *Se non è vero, è ben trovato.* If it is not true, it deserves to be true.

If, then, the statement is true, it is a coincidence of a more complex form than the simple triangle case ($n = 3$) and we should look for a fairly simple explanation. If the statement is false for $n = 4$, there are still, obviously, special instances of it which are true, e.g., the regular tetrahedra, and we would be inclined to deem interesting that class of tetrahedra for which it is

true. In either case, then, natural curiosity and aesthetics would start from the sense of coincidence and force an appropriate theory to come into being.

(See [**2**].)

Example 6. A number of years ago, in the April issue of the *Scientific American*, Martin Gardner, who writes a regular column of mathematical recreation, ran a special "April Fool's" column. This was a collection of scientific hoaxes. One statement in particular caught the eyes of people who enjoy doing long computations:

"The number $N = e^{\pi\sqrt{163}}$ is an integer."

Now this statement is a shocker. If it is true, we would have a "fairly simple" real relationship between π and e. In view of all we know or have heard about such numbers, this does not seem likely. In view of the known complex relationship $e^{\pi\sqrt{-1}} = -1$, however, there might, perhaps, be a residue of doubt in our minds. N just might be an integer.

Well, let's go to the computer. But notice that a little hand-held computer is useless, for $e \approx 2.7$, $\pi \approx 3.1$, $\sqrt{163} \approx 12.8$, so that $N \approx e^{40} \approx 10^{17}$. Since hand computers carry 10 to 12 figures, we have got to go to a computer where we can work in multiple precision. Numbers such as $e, \pi, \sqrt{163}$ have been computed out to hundreds, if not thousands, of figures and have been published, but we had best write our own superaccurate programs for these numbers. At any rate, let us shoot for 20 figures.

A 20-figure computation yields $N = 262{,}53741{,}26407{,}68743.99$. Hmm. What have we here? The fractional part .99 is tantalizing. Let's go for 25 figures. The computer grinds away (requiring a time which is probably proportional to the cube of the number of figures desired), but finally prints out

$$N = 262{,}53741{,}26407{,}68743.99999\ 99.$$

The mystery thickens. Is it really possible that N is an integer? Of course one can't prove it by this kind of numerical computation, for however long a string of 9's were produced there would be some residue of doubt. But we could *disprove it*; so let's push the button once again and give the computer a really good whirl.

$$N = 262{,}53741{,}26407{,}68743.99999\ 99999\ 99250.$$

Well, that blows it. N isn't an integer after all. But notice: it differs from an integer by around 10^{-12}.

Now of course anyone well versed in the theory of transcendentals could have told us that $e^{\pi\sqrt{163}}$ is not an integer. It is what we think it should be: a transcendental number. This can be deduced as a consequence, for example, of the Gelfond-Schneider Theorem (1932), which tells us that if a is an algebraic number and is neither 0 nor 1, and if b is an algebraic irrational, then a^b is a transcendental.

Now $N = e^{\pi\sqrt{163}} = e^{-\pi i\sqrt{-163}} = (-1)^{\sqrt{-163}}$, so that by Gelfond-Schneider, N is transcendental. But the equation $N = \text{integer} + O(10^{-12})$ is remarkable, and surely requires some sort of an explanation. You just don't slap e's and π's and square roots around and expect to get integers to within 10^{-12}. (Forget obvious identities such as $\pi/\pi + e/e = 2$.) On the other hand, a 12-figure coincidence doesn't seem excessively long (certainly not as good as the 26-figure coincidence in Example 3); so one would expect that the explanation would be fairly complicated. And complicated it is.

The writer was first shown this mathematical amusement by Morris Newman in 1952 when he joined the National Bureau of Standards. Newman was a student of Hans Rademacher and had just computed N to high precision on one of the first electronic computers in the world: the SEAC. It is not unreasonable to suppose that the explanation of the coincidence is to be found in the kind of mathematics that Rademacher delighted in and was most skillful in: analytic and

Numerical agreement is reported to 43 places. Does $I = L$?

Query: Was numerical agreement obtained as a result of a computation or inferred from a proof of $I = L$?

(See [**16**].)

Example 7. The Riemann Hypothesis is currently the most notorious of unsolved problems of analysis, if not of the whole of mathematics. Riemann's zeta function is defined for $\mathrm{Re}\, z > 1$ by means of the Dirichlet Series

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

and is extended to the rest of the complex plane (where possible) by analytic continuation. The Riemann Hypothesis is that all the complex roots of the zeta function have real part equal to $\frac{1}{2}$. Many interesting consequences of this theorem are known, many equivalent formulations are known, but thus far a proof or a disproof has eluded investigators.

Many years ago, G. H. Hardy proved that infinitely many roots of the zeta function have real part equal to $\frac{1}{2}$. By laborious computations and laborious analyses of truncation errors, it has been shown that the first 70,000,000 roots all have the real parts equal to $\frac{1}{2}$. Coincidence? Or is this sufficient evidence to convince one of the truth of the Riemann Hypothesis?

Some authorities say no. They point out that in the theory of the zeta function, as well as in prime number theory, functions whose growth is as slow as that of $\log\log x$ occur often, and since $\log\log x = 10$ if x is around $10^{10{,}000}$, of what significance is a paltry 70,000,000? There is even a case in the literature of primes when a prevailing tendency was proved by Littlewood to cease prevailing when $n > 10^{10^{10^{34}}}$ (Skewes's number, since lowered.)

But Good and Churchhouse have latched onto another "coincidence." The Möbius function $\mu(n)$ is defined as follows:

$$\mu(n) = +1 \text{ if } n \text{ has an even number of distinct prime factors}$$
$$\mu(n) = -1 \text{ if } n \text{ has an odd number of distinct prime factors}$$
$$\mu(n) = 0 \quad \text{if } n \text{ has a repeated prime factor.}$$

The "visual" evidence shows that $\mu(n)$ behaves in an erratic fashion. One can show without too much difficulty that the probability

$$\text{that } \mu(n) = +1 \text{ is } 3/\pi^2$$
$$\text{that } \mu(n) = -1 \text{ is } 3/\pi^2$$
$$\text{that } \mu(n) = 0 \text{ is } 1 - 6/\pi^2.$$

Now, the Strong law of large numbers in probability theory tells us that if μ_n is a random variable selected N times with the probabilities just listed, then

$$\sum_{n=1}^{N} \mu_n < CN^{1/2+\epsilon}$$

with probability one. But it has been known for some time that the Riemann Hypothesis is equivalent to the inequality

$$\sum_{1}^{N} \mu(n) \leqslant CN^{1/2+\epsilon}.$$

Hence, if we could somehow identify N selections at random of μ from the distribution above with the first N deterministic values $\mu(1), \ldots, \mu(N)$, then the Riemann Hypothesis would be proved. To check numerically whether this identification is valid, Good and Churchhouse computed $\mu(n)$ for

$n \leqslant 33{,}000{,}000$. The number of n's for which $\mu(n) = 0$ was 12,938,407. Now

$$33{,}000{,}000(1 - 6/\pi^2) = 12{,}938{,}405.6.$$

Here we have 8-figure accuracy. Incredible!

Despite these coincidences, authorities have asserted that there is still but paltry evidence for believing in the truth of the Riemann Hypothesis. But it seems to me that this great coincidence shows either (1) the Riemann Hypothesis is true or (2) you have got a very difficult theorem on your hands to explain away (the 8-figure accuracy). And it is psychologically and almost morally incumbent upon you to do so.

(See [5], [**8**], [**11**], [**13**].)

Example 8. *Pappus's Theorem*; *Program Verification*. One of the principles that is drummed into every aspiring young mathematician is that one cannot prove a general theorem by merely proving special instances of it. One must present a proof that embraces all cases. Not infrequently does a student get his knuckles rapped for dealing only with a special case. (You assumed the triangle has a right angle; you assumed that α was real, but it may be complex; and so on.) Yet this principle contains an aspect which is palpably wrong. In the first place, the general theorem has been inferred, like as not, from several special cases. Second, if the special case is sufficiently random or "odd-ball," then its truth implies—with high probability—the general truth and, as we shall see, for "sufficiently odd-ball" cases, it implies the exact truth.

Consider, as an example, a classical theorem of fourth-century mathematics proved by Pappus. In Fig. 1, the lines l_1 and l_2 are selected at will and on these points $P_1, \ldots, P_6$ are selected at will. They are then cross-connected as indicated.

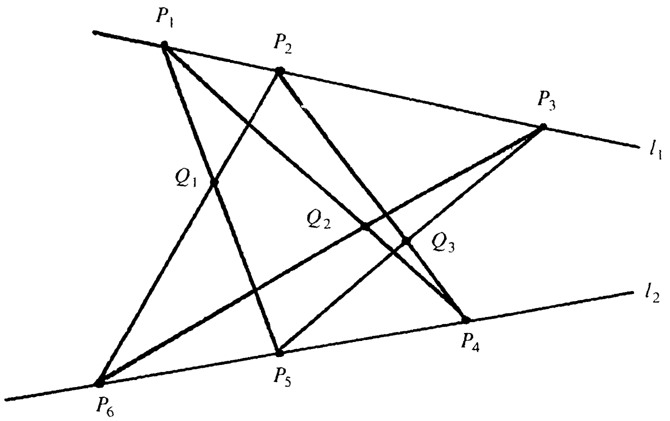


FIG. 1

The theorem now asserts that the points Q_1, Q_2, Q_3 will always be collinear. This theorem may very well have been found experimentally by fooling around with a ruler and pencil. Its ingredients are of the very simplest.

It should be observed that Pappus's proof and updated versions of it to be found in elementary projective geometry embody an element of great ingenuity. Also, this theorem lies at the heart of abstract projective geometry over arbitrary algebraic structures insofar as it is equivalent to the commutativity of the elements of the structure.

The verification of Pappus's theorem in any particular rational numerical case can be carried out completely in the field of rational numbers. Hence it can be carried out with exact precision on a digital computer. The assertion that Q_1, Q_2, Q_3 are collinear is equivalent to the statement

that

$$D = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} = 0$$

where $Q_i = (x_i, y_i)$. Now, beginning from integers of smallish size for the coordinates of $P_1, \ldots, P_6$, the arithmetic mixing process intrinsic in the geometric figure leads to very large integers for the numerator and denominator of the points Q_i. Therefore the discovery that $D = 0$ becomes an incredible arithmetic identity unless, of course, one believes that the theorem is true in general.

If now one allows the points $P_1, \ldots, P_6$ to have irrational coordinates, then the numerical verification of the identity $D = 0$ is, in the strict sense, impossible on a computer. However, it can be done by setting up the irrational quantity as an abstract variable, to be processed as a variable and not numerically, and this is tantamount to a rigorous proof.

In a similar way, consider a program P residing in a digital computer. Let us suppose it is such that given an input value x, the computer prints out the value $y = P(x)$. We have a certain ideal program $\tilde{P}$ in mind and we should like to inquire whether $P \equiv \tilde{P}$. In most instances, it is simply out of the question to test whether $P(x) = \tilde{P}(x)$ for all the possible input values x.

To limit the problem, suppose we know a priori that both P and $\tilde{P}$ are selected from a limited family of possible programs. Say they are both polynomials in a single variable x of degree $\leqslant n$. Then, from our theory of polynomials, we can assert that if $x_1, \ldots, x_{n+1}$ are n distinct inputs and if $P(x_i) = \tilde{P}(x_i)$ $i = 1, \ldots, n+1$, then $P \equiv \tilde{P}$. Thus a limited testing implies complete identity. We can assert more.

If we test out the program with a transcendental element x^* (which, in the numerical sense, is inaccessible to the computer), then $P(x^*) = \tilde{P}(x^*)$ implies $P \equiv \tilde{P}$. Therefore, one special (but sufficiently odd-ball) case implies complete identity.

But we can assert more. Suppose that P and $\tilde{P}$ are extracted from the class of *computable functions*. Then a simple countability argument provides us with a value x^* such that $P(x^*) = \tilde{P}(x^*)$ implies $P \equiv \tilde{P}$. Thus we have in this statement a theory of coincidences wherein a single coincidence of sufficient intensity implies complete identity.

(See [**1**], [**3**], [**4**], [**6**], [**7**], [**12**], [**17**].)

Example 9. This example is of a different order. The Laplace equation

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

plays a role in the theory of gravitational attraction, in electro- and magneto-statics, in fluid flow, in the theory of elasticity, etc. Coincidence, or part of a Grand Order? Again, to inject a personal note, when I first learned this, it had the strong flavor of a coincidence, and my teacher and my textbook certainly emphasized this aspect as well as the fact that one wouldn't have to learn totally different theorems. The solution of problems of fluid flow by means of electrostatic analog devices seemed to confirm the miracle.

Today the books are more sophisticated and attempt to unify the whole by pointing out that all those physical phenomena are subject to the $1/r$ force law, to similar laws of invariance, of conservation, and of the operation of the maximum principle.

To restore a sense of coincidence here, consider a different aspect of the equations of mathematical physics. The relationship between real, two-dimensional potential theory and the theory of analytic functions of a complex variable is one of the great glories of mathematics. Many novices anticipate that there must be a corresponding relationship between higher dimensional potential theory and a theory of functions of a hypercomplex variable.

This is not true in the simplistic sense. But the drive toward preserving the coincidence has brought forth theories, e.g., application of quaternions or Bergman's theory of integral operators of several complex variables, which in some degree can be regarded as substitutes. (See [**10**].)

But enough of examples: It is time now that we drew some inferences from them.

From time to time mathematicians perceive certain similarities of form which elicit an element of surprise. Such a similarity may be called a coincidence. The surprise calls for an explanation. The explanation, if it is forthcoming, serves partially to kill the surprise.

The existence of the coincidence implies the existence of an explanation. If the coincidence is of a high degree of improbability, then there is more to explain and the explanation will be easier in the sense that it involves a more easily accessible theory. If the coincidence is only of a medium order of improbability, the explanation will be more difficult.

A Platonic philosophy of mathematics might say that there are no coincidences in mathematics because all is ordained. In the words of Alexander Pope (*Essay on Man*):

All nature is but art unknown to thee
All chance, direction which thou can'st not see.

But for the working mathematician, coincidence exists. He feels it, he identifies it, he uses it as an inductive and constructive element. He pursues its implication along certain lines. To some extent, he even brings it about.

Acknowledgment. I wish to thank Professors Reuben Hersh, Lawrence D. Kugler, and Richard A. Vitale, who read a preliminary version of this paper and made many valuable comments.

References

1. G. J. Chaitin and J. T. Schwartz, A note on Monte Carlo primality tests and algorithmic information theory, Comm. Pure Appl. Math., 31 (1978) 521–527.

2. H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Mathematical Library, no. 19, MAA, 1967.

3. E. Cerutti and P. J. Davis, FORMAC meets Pappus: Some observations on elementary analytic geometry by computer, this MONTHLY, 76 (1969) 895–905.

4. P. J. Davis, Proof, completeness, transcendentals, and sampling, J. Assoc. Comput. Mach., 24 (1977) 298–310.

5. P. J. Davis and R. Hersh, The Mathematical Experience, Birkhäuser, Boston, 1980.

6. R. A. DeMillo and R. J. Lipton, A probabilistic remark on algebraic program testing, Inf. Processing Letters, 7, no. 4 (1978) 193–195.

7. R. A. DeMillo, R. J. Lipton, and A. J. Perlis, Social processes of proofs of theorems and programs, GIT-ICT-78/04 (Georgia Institute of Technology).

8. H. E. Edwards, Riemann's Zeta Function, Academic Press, New York, 1974.

9. K. O. Friedrichs, From Pythagoras to Einstein, New Mathematical Library, no. 19, MAA, 1965, p. 36.

10. R. P. Gilbert, Function Theoretic Methods in Partial Differential Equations, Academic Press, New York, 1969.

11. I. J. Good and R. F. Churchhouse, Riemann hypothesis and pseudorandom features of the Möbius sequence, Math. Comp., 22 (1968) 857–861.

12. W. E. Howden, Algebraic Program Testing, Comp. Sci. Tech. Rep. no. 14, University of California-San Diego, La Jolla, Calif., November 1976.

13. M. Kac, Statistical Independence in Probability, Analysis and Number Theory, Carus Monograph 12, MAA, 1959.

14. M. Newman, personal communications.

15. H. O. Pollak, R. L. Graham, personal communications.

16. H. P. Robinson, Problem 80–7, SIAM Review, 22 (1980) 229.

17. J. H. Rowland and P. J. Davis, On the use of transcendentals for program testing, J. Assoc. Comput. Mach. (to appear).

18. D. Shanks, Incredible identities, Fibonacci Quarterly, 12 (1974) 271, 281.

19. Harold M. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.

20. ______, On a gap in a theorem of Heegner, J. Number Theory, 1 (1969) 16–27.

21. H. Weber, Lehrbuch der Algebra, Elliptische Funktionen und Algebraische Zahlen, vol. 3, 1908; reprinted, Chelsea, New York, 1961.

ELEMENTARY CHARACTERIZATION OF CLASSICAL MINIMA

W. HRUSA
Division of Applied Mathematics, Brown University, Providence, RI 02912

J. L. TROUTMAN
Department of Mathematics, Syracuse University, Syracuse, NY 13210

1. Introduction. It is a curious feature of the historical development of a theory for optimization that attention was first directed toward finding necessary conditions which then often were regarded (tacitly) as being sufficient. Moreover, a century after the need for separate consideration of sufficiency was made apparent (notably through the efforts of Weierstrass), textbooks still introduce and often define the subject in terms of its necessary conditions. This is largely the case with the classical calculus of variations, which is dominated by the study of the necessary conditions supplied through the Euler-Lagrange equations. Sufficiency is then approached through the much more technical and sophisticated device of field theory (which falls outside the scope of many introductory treatments and is consequently ignored).

It is therefore surprising that several optima of classical interest usually treated through the calculus of variations can be characterized directly by analysis much more elementary than the Euler-Lagrange theory. We shall substantiate this assertion by using little more than the Cauchy-Schwarz vector inequality to provide direct access to the solutions of several well-known problems, including certain of those associated with the brachistochrone and the minimal surface.

We begin with two almost trivial examples from the classical literature [**2**, pp. 178–180, 201], namely those of determining geodesics in space and on the surface of a sphere.

To find the shortest space curve joining fixed points $P_0(=0)$ and P_1, we must minimize

$$\int_0^1 |P'(t)|\,dt$$

over smooth vector-valued functions $P(t)$ subject to $P(0)=0$ and $P(1)=P_1$. We have the following inequality:

$$\int_0^1 |P'(t)|\,dt \geqslant \left|\int_0^1 P'(t)\,dt\right| = |P_1|$$

with equality if $|P'(t)| = |P_1|$. Thus as expected, the line segment defined by $P(t) = P_1 t$, $0 \leqslant t \leqslant 1$, gives the shortest distance between $P_0 = 0$ and P_1.

The same technique can be used to find the smooth geodesics on a sphere of radius a. Assuming, as we may, that a smooth curve originates at the north pole ($\phi = 0$) and can be represented as the graph of a function $\theta = \theta(\phi)$, $0 < \phi \leqslant \phi_1 \leqslant \pi$, with $\theta(\phi_1) = \theta_1$, in spherical surface coordinates (a,θ,ϕ), then its length is given by

$$L(\theta) = a\int_0^{\phi_1}\left[1 + (\theta'\sin\phi)^2\right]^{1/2} d\phi,$$

and clearly

$$L(\theta) \geqslant a\int_0^{\phi_1} d\phi = a\phi_1$$

with equality iff $\theta'(\phi) = 0 \Leftrightarrow \theta(\phi) = \text{constant} = \theta_1$. Thus the geodesics on a sphere are precisely the shorter arcs of great circles, as is well known.

William Hrusa is a graduate student at Brown University, where he received a master's degree in applied mathematics in 1979.

John L. Troutman, Associate Professor of Mathematics at Syracuse University, received advanced degrees in applied mechanics from Virginia Polytechnic Institute and Stanford University and a doctorate in mathematics from Stanford. His research interests are primarily in the fields of complex analysis and potential theory.—*Editors*

This method of finding a lower bound for the integral is limited in its applications. For example, its success in the previous problem depended upon presence of the free boundary condition at $\phi = 0$. A more fruitful approach is suggested by a standard reformulation.

2. A Simple Proposition. Given $x_0, x_1, y_0, y_1 \in \mathbb{R}$ with $x_0 < x_1$, for $k = 1,2$, let $C^k[x_0,x_1]$ denote the subspace of $C[x_0,x_1]$ consisting of functions with continuous derivatives of order k on (x_0,x_1) admitting continuous extensions to $[x_0,x_1]$ and set

$$\mathfrak{D} = \{y \in C^1[x_0,x_1] : y(x_j) = y_j, j = 0,1\}$$

and

$$\mathfrak{D}_0 = \{v \in C^1[x_0,x_1] : v(x_j) = 0, j = 0,1\}.$$

We shall be concerned with the classical problem of minimizing on $\mathfrak{D}$ the function

$$J(y) \stackrel{\text{def}}{=} \int_{x_0}^{x_1} F(x,y(x),y'(x))\,dx$$

where the real valued function F is assumed to be such that $F(x,y(x),y'(x))$ is continuous (or at least integrable) on $[x_0,x_1]$ for each $y \in \mathfrak{D}$.

The following is obvious: $y \in \mathfrak{D}$ minimizes J on $\mathfrak{D}$ iff $\Delta J(y;v) \stackrel{\text{def}}{=} J(y+v) - J(y) \geqslant 0$ $\forall v \in \mathfrak{D}_0$; if, in addition, $\Delta J(y;v) = 0 \Rightarrow v = 0$, then y provides the *unique* minimizing function for J on $\mathfrak{D}$. If we can find a suitable lower bound for ΔJ on $\mathfrak{D} \times \mathfrak{D}_0$, we may then be able to characterize by inspection a minimizing y for which this lower bound vanishes. The remaining examples yield to this approach.

For most of these examples, we shall need to consider only one class of functions J.

PROPOSITION. *If*

$$J(y) = \int_{x_0}^{x_1} \alpha(x)\left[1 + (\beta y)^2(x) + (\gamma y')^2(x)\right]^{1/2} dx$$

where $\alpha,\beta,\gamma \in C[x_0,x_1]$, and both α and $\beta^2 + \gamma^2$ are positive on (x_0,x_1), then with the above definitions of $\mathfrak{D}$ and $\mathfrak{D}_0$,

$$\Delta J(y;v) \geqslant I(y;v) \stackrel{\text{def}}{=} \int_{x_0}^{x_1} \alpha\left[\frac{\beta^2 yv + \gamma^2 y'v'}{(1 + \beta^2 y^2 + \gamma^2 y'^2)^{1/2}}\right](x)\,dx, \qquad \forall y \in \mathfrak{D}, \quad v \in \mathfrak{D}_0, \tag{1}$$

with equality at y iff $v = 0$. Thus if $\exists y \in \mathfrak{D}$ for which $I(y;v) = 0$, $\forall v \in \mathfrak{D}_0$, then y provides the unique minimizing function for J on $\mathfrak{D}$.

Proof. For $y \in \mathfrak{D}, v \in \mathfrak{D}_0$: the standard Cauchy-Schwarz vector inequality in the form

$$|A| = |B|^{-1}(A \cdot B)\sec(A,B) \geqslant |B|^{-1}(A \cdot B),$$

with equality iff A and B are co-directed, shows for the vectors $A = (1,\beta(y+v),\gamma(y'+v'))$ and $B = (1,\beta y,\gamma y')$ that at each $x \in (x_0,x_1)$:

$$\begin{aligned}&\left[1 + \beta^2(y+v)^2 + \gamma^2(y'+v')^2\right]^{1/2}\\ &\qquad \geqslant \left[1 + \beta^2 y^2 + \gamma^2 y'^2\right]^{-1/2}\left[1 + \beta^2 y(y+v) + \gamma^2 y'(y'+v')\right]\end{aligned}$$

or upon utilizing the positivity of α,

$$\begin{aligned}&\alpha\left[1 + \beta^2(y+v)^2 + \gamma^2(y'+v')^2\right]^{1/2} - \alpha\left[1 + \beta^2 y^2 + \gamma^2 y'^2\right]^{1/2}\\ &\qquad \geqslant \frac{\alpha(\beta^2 yv + \gamma^2 y'v')}{\left[1 + \beta^2 y^2 + \gamma^2 y'^2\right]^{1/2}}\end{aligned} \tag{2}$$

with equality at x only if $v(x)v'(x)=0$ since by hypothesis $\alpha(x)$ and either $\beta(x)$ or $\gamma(x)$ is nonvanishing. Integrating (2) over $[x_0,x_1]$ and recalling the definition of ΔJ for this specific choice of J gives (1). Clearly $v=0$ implies equality in (1). Conversely, equality of the integrals represented in (1), in the presence of (2), implies (through continuity) equality in (2) on (x_0,x_1), and hence the vanishing there of $(v^2)'=2vv'$. Thus, $v^2(x)=\text{const.}=v^2(x_0)=0$, so that $v=0$. ■

In effect, the Cauchy-Schwarz inequality as used above establishes the strict convexity of F in its last two arguments; an analogous proposition holds for such F, but convexity considerations are not essential to this presentation.

REMARK 1. The proposition also holds if the function α is positive and integrable over $[x_0,x_1]$.

REMARK 2. If $\exists y\in\mathfrak{D}$ for which $I(y;v)=0$ (as in the preceding proposition) $\forall v\in\mathfrak{D}_0$, then standard approximation arguments show that also $I(y;\hat{v})=0$, $\forall\hat{v}\in\hat{\mathfrak{D}}_0$, where

$$\hat{\mathfrak{D}}_0=\left\{\hat{v}\in C[x_0,x_1]:\hat{v}\text{ is piecewise }C^1\text{ on }[x_0,x_1]\text{ and }\hat{v}(x_j)=0, j=0,1\right\}.$$

J is also defined on

$$\hat{\mathfrak{D}}=\left\{\hat{y}\in C[x_0,x_1]:\hat{y}\text{ is piecewise }C^1\text{ on }[x_0,x_1]\text{ and }\hat{y}(x_j)=y_j, j=0,1\right\}$$

and the preceding proof extends to show that $\Delta J(y;\hat{v})\geqslant I(y;\hat{v})=0$, $\forall\hat{v}\in\hat{\mathfrak{D}}_0$, with equality iff $\hat{v}=0$ so that this *same* $y\in\mathfrak{D}$ is the *unique* minimizing function for J on $\hat{\mathfrak{D}}$. [Indeed, for $v=\hat{v}$, although the integrands on both sides of (2) each have the same points of discontinuity as $v'(=\hat{v}')$, the inequality is preserved at the limiting values from either side together with conditions for its sharpness. In particular, equality of the corresponding integrals is possible only when $\hat{v}\hat{v}'=0$ *where defined*. Thus $\hat{v}^2$ is continuous and piecewise constant on $[x_0,x_1]$ and so must vanish with $\hat{v}^2(x_0)$.]

REMARK 3. Observe that with

$$F(x,y,y')=\alpha(x)\left[1+(\beta(x)y)^2+(\gamma(x)y')^2\right]^{1/2},$$
$$I(y;v)=\int_{x_0}^{x_1}\left[F_y(x,y(x),y'(x))v(x)+F_{y'}(x,y(x),y'(x))v'(x)\right]dx.$$

Since we are interested only in sufficiency, we may seek the minimizing $y\in\mathfrak{D}\cap C^2[x_0,x_1]$. Thus if $\alpha,\beta,\gamma\in C^1[x_0,x_1]$ we may integrate the second term by parts to get

$$I(y;v)=\int_{x_0}^{x_1}\left[F_y(x,y(x),y'(x))-\frac{d}{dx}F_{y'}(x,y(x),y'(x))\right]v(x)\,dx,\qquad\forall v\in\mathfrak{D}_0,$$

and see by inspection that any $y\in\mathfrak{D}\cap C^2[x_0,x_1]$ which satisfies the Euler-Lagrange differential equation

$$F_y(x,y(x),y'(x))-\frac{d}{dx}F_{y'}(x,y(x),y'(x))=0$$

must provide uniquely the minimum for J on $\mathfrak{D}$ (and on $\hat{\mathfrak{D}}$ as in Remark 2). Of course we have no assurance that this equation has a solution in $\mathfrak{D}\cap C^2[x_0,x_1]$.

REMARK 4. In the simple case that $\beta\equiv 0$, it is seen *by inspection* that any $y\in\mathfrak{D}$ which makes $\alpha\gamma^2y'/(1+\gamma^2y'^2)^{1/2}$ constant on (x_0,x_1) will provide the unique minimizing function for J on $\mathfrak{D}$ (and on $\hat{\mathfrak{D}}$). In particular, when in addition α and γ are constant, then the linear function defined on $[x_0,x_1]$ by $y(x)=[(y_1-y_0)/(x_1-x_0)(x-x_0)]+y_0$ is in $\mathfrak{D}$ and *must* provide the unique minimizing function for J on $\mathfrak{D}$ (and on $\hat{\mathfrak{D}}$).

REMARK 5. A similar analysis can be made when

$$J_R(y) = \int_{x_0}^{x_1} \left[(\beta y)^2(x) + (\gamma y')^2(x)\right]^{1/2} dx$$

if β,γ are continuous and nonvanishing on $[x_0,x_1]$ and $\mathfrak{D}$ is replaced by

$$\mathfrak{D}_R = \left\{y \in \mathfrak{D} : y^2 + y'^2 > 0 \text{ on } [x_0,x_1]\right\}.$$

In this case,

$$\Delta J_R(y;v) \geqslant I_R(y;v) = \int_{x_0}^{x_1} \left[\frac{\beta^2 yv + \gamma^2 y'v'}{\left((\beta y)^2 + (\gamma y')^2\right)^{1/2}}\right](x)\,dx, \qquad \forall y \in \mathfrak{D}_R, v \in \mathfrak{D}_0,$$

with equality at y only if $vy' = v'y$ on (x_0,x_1); i.e., only if $v =$ const. y and for $v \in \mathfrak{D}_0$, this implies that $v = 0$ (unless $y_0 = y_1 = 0$). Thus if either y_0 or y_1 is nonzero, then again any $y \in \mathfrak{D}_R$ which makes $I_R(y;v) = 0$, $\forall v \in \mathfrak{D}_0$, must provide the unique minimizing function for J_R on $\mathfrak{D}_R$ (and on $\hat{\mathfrak{D}}_R = \{\hat{y} \in \hat{\mathfrak{D}} : \hat{y}^2 + \hat{y}'^2 > 0 \text{ on } [x_0,x_1]\}$ as in Remark 2).

3. Applications.

Geodesic Problems. 1. To find (nontrivial) geodesics on a right circular cylinder of radius one unit joining points $P_0 = (\theta_0,y_0)$ and $P_1 = (\theta_1,y_1)$ (in cylindrical coordinates $(1,\theta,y)$) we suppose as we may that $0 < \theta_1 - \theta_0 \leqslant \pi$ and consider the problem of minimizing

$$L(y) = \int_{\theta_0}^{\theta_1} \left(1 + y'(\theta)^2\right)^{1/2} d\theta$$

on $\mathfrak{D} = \{y \in C^1[\theta_0,\theta_1] : y(\theta_j) = y_j, j = 0,1\}$. With an obvious replacement of variables, L is seen to be a function of the simplest type discussed in Remark 4 with $\alpha(x) = \gamma(x) \equiv 1$ and $\beta(x) \equiv 0$.

Thus the circular helix with the linear representation

$$y(\theta) = \frac{y_1 - y_0}{\theta_1 - \theta_0}(\theta - \theta_0) + y_0, \qquad \theta \in [\theta_0,\theta_1],$$

is the (unique) curve of shortest length among all those representable by functions $y \in \mathfrak{D}$ (or in $\hat{\mathfrak{D}}$, the corresponding space of piecewise C^1 functions discussed in Remark 2).

2. To consider the (nontrivial) smooth geodesics between points P_0 and P_1 on a right circular cone of apex angle 2ϕ, we suppose that the cone has its apex at the origin of a Cartesian coordinate system and that its axis of symmetry is the positive z-axis. Then a point on the conical surface represented by (r,θ,z) in cylindrical coordinates has the rectangular coordinates $(x,y,z) = (r\cos\theta, r\sin\theta, ar)$, where $a = \cot\phi$, and is thus specified by coordinates r,θ. In particular, $P_k = (r_k,\theta_k), k = 0,1$, and we can suppose that $0 < r_0 \leqslant r_1$, $\theta_0 = 0$, and $0 < \theta_1 \leqslant \pi$.

The most interesting curves amenable to our methods are those representable as graphs of functions in

$$\mathfrak{D}_R = \left\{r \in C^1[0,\theta_1] : r(0) = r_0, r(\theta_1) = r_1 \text{ and } r^2 + r'^2 > 0 \text{ on } [0,\theta_1]\right\}$$

(see Remark 5). Each $r \in \mathfrak{D}_R$ represents a curve of length

$$L(r) = \int_0^{\theta_1} \left[r(\theta)^2 + b^2 r'(\theta)^2\right]^{1/2} d\theta$$

(where $b = \sqrt{1 + a^2}$) which by Remarks 3 and 5 would be minimized (uniquely) on $\mathfrak{D}_R$ by any $r \in \mathfrak{D}_R \cap C^2[0,\theta_1]$ which satisfies the Euler-Lagrange differential equation

$$\frac{d}{d\theta} \frac{b^2 r'(\theta)}{\left[r(\theta)^2 + b^2 r'(\theta)^2\right]^{1/2}} = \frac{r(\theta)}{\left[r(\theta)^2 + b^2 r'(\theta)^2\right]^{1/2}}.$$

With the trigonometric substitution $b(r'/r) = \tan\psi$, the differential equation simplifies to

$$b(d/d\theta)\sin\psi(\theta) = \cos\psi(\theta)$$

which has as the obvious solution $\psi'(\theta) = b^{-1} = \sin\phi$ resulting in $b(r'/r) = \tan(b^{-1}\theta + c)$, or

$$r(\theta) = c_1\sec(b^{-1}\theta + c) \tag{3}$$

for constants c,c_1 to be chosen if possible so that the boundary conditions can be satisfied. Now, $0 < b^{-1}\theta_1 < \pi$, and we should choose c so that

$$\frac{\cos(b^{-1}\theta_1 + c)}{\cos c} = \frac{r_0}{r_1} = \rho,$$

say, or $\tan c = \cot b^{-1}\theta_1 - \rho\csc b^{-1}\theta_1$, and this is always possible. With this c, we need only take $c_1 = r_0\cos c$ to satisfy both boundary conditions.

Thus the curve defined by (3) for appropriately determined constants c,c_1 has (uniquely) the minimum length among all those represented by functions r in $\mathfrak{D}_R$ (or in $\hat{\mathfrak{D}}_R$ the corresponding class of piecewise C^1 functions discussed in Remark 2). In particular, we see that when $r_0 = r_1$, the minimum length is *not* given by the associated circumference as might be conjectured.

Time of Transit Problems. 1. The classical problem most influential in the development of the calculus of variations is that of the brachistochrone, which was first considered by Galileo c. 1635 and was reactivated by the Bernoullis, John and James, in 1696. It consists of characterizing the shape of a planar wire joining a pair of horizontally and vertically displaced points along which a bead will slide from rest in minimum time under the action of a constant gravitational force. If we utilize a coordinate system in which the *x-axis* is directed with the force, then any smooth curve from the origin to a lower point (x_1,y_1) which admits representation as the graph of a function $y \in C^1[0,x_1]$, requires the transit time

$$T(y) = \int_0^{x_1}(2gx)^{-1/2}\left(1 + y'(x)^2\right)^{1/2}dx$$

([**6**, p. 8]) and we wish to minimize T on

$$\mathfrak{D} = \left\{y \in C^1[0,x_1]: y(0) = 0, y(x_1) = y_1\right\}.$$

This function T is covered by Remark 1 with $\alpha(x) = (2gx)^{-1/2}$ on $[0,x_1]$ and Remark 4 with $\beta(x) \equiv 0$ and $\gamma(x) \equiv 1$. Thus we should seek a $y \in \mathfrak{D}$ for which

$$x^{-1/2}y'(1 + y'^2)^{-1/2} = \text{const.} = \sqrt{2}\,c^{-1}, \tag{4}$$

say. This equation can be integrated after the introduction of a new independent variable θ defined by $x(\theta) = c^2(1 - \cos\theta) = 2c^2\sin^2(\theta/2)$ on $[0,\pi]$, and there results the equation $y(\theta) = c^2(\theta - \sin\theta)$ where we have incorporated the condition $y(0) = 0$.

The resulting pair of equations for $x(\theta)$ and $y(\theta)$ are well known to be the parametric equations of a cycloid with a cusp at the origin. It is known that for a given $x_1 > 0, y_1 > 0$ there is a (unique) cycloid of this type joining the origin to (x_1,y_1) ([**1**, p. 55]). Moreover, when $y_1/x_1 \leq \pi/2$, then (x_1,y_1) lies on the first half of the arch of the cycloid joining it to the origin, and when $y_1/x_1 < \pi/2$, the associated cycloid admits representation as the graph of a function $y \in \mathfrak{D}$; i.e., the defining function for $x(\theta)$ is invertible and the resulting composite $y(\theta(x))$ provides an element of $\mathfrak{D}$.

Thus when $y_1/x_1 < \pi/2$ the cycloid (so represented) has (uniquely) the least time of descent among all curves representable as the graphs of functions $y \in \mathfrak{D}$ (or in $\hat{\mathfrak{D}}$, the corresponding class of piecewise C^1 curves discussed in Remark 2).

It is possible to extend our analysis to the case $y_1/x_1 = \pi/2$ by a limiting argument. However, when $y_1/x_1 > \pi/2$, we can draw no conclusions from this analysis alone. Nevertheless, it seems remarkable that even this partial solution to the problem of the brachistochrone is obtainable through the elementary methods employed here.

2. We consider next a simple case of the Zermelo navigation problem. It consists of determin-

ing a smooth path joining fixed points on opposite banks of a river with varying current along which the transit time will be minimized for a boat which travels at constant unit speed with respect to the water. We suppose that the river banks are parallel and utilize a coordinate system in which the y-axis coincides with one bank and the line $x = x_1$ with the other. We also assume that the current e is always directed downstream, admits the prescription $e = e(x)$ (continuous) on $[0,x_1]$, and satisfies $0 \leqslant e(x) < 1$ (so that each point on the opposite bank is attainable). Any smooth path from the origin to a downstream point (x_1,y_1) which admits representation as the graph of a function $y \in C^1[0,x_1]$ requires the transit time

$$T(y) = \int_0^{x_1}\Big[\alpha(x)\big(1 + (\alpha y')^2(x)\big)^{1/2} - (\alpha^2 e y')(x)\Big]\,dx$$

where $\alpha(x) \equiv (1 - e^2(x))^{-1/2}$([5, p. 13]). Thus we wish to minimize T on

$$\mathfrak{D} = \{y \in C^1[0,x_1]: y(0) = 0, y(x_1) = y_1\}.$$

Applying the Cauchy-Schwarz vector inequality to the first term in the brackets as in the proof of the Proposition, we get with the obvious definitions that

$$T(y + v) - T(y) = \Delta T(y;v) \geqslant \int_0^{x_1}\Big\{(\alpha^3 y')(x)\big[1 + (\alpha y')^2(x)\big]^{-1/2} - (\alpha^2 e)(x)\Big\}v'(x)\,dx,$$
$$\forall y \in \mathfrak{D}, \quad v \in \mathfrak{D}_0 \qquad (5)$$

with equality *if and only if* $v = 0$. The integral vanishes when the term in braces is constant $= c$, say, and its vanishing for some $y \in \mathfrak{D}$ determines that y is the unique minimizing function for T on $\mathfrak{D}$ by the same considerations as before.

It is straightforward to verify that the function

$$y(x) = \int_0^x (e + c\alpha^{-2})[1 - 2ce - c^2\alpha^{-2}]^{-1/2}(t)\,dt$$

does indeed satisfy the differential requirement for arbitrary choice of the constant c which makes the square bracket positive, and clearly $y(0) = 0$. However, it is not always possible to find c to satisfy the remaining condition $y(x_1) = y_1$; for example, the linear profile $e(x) = (1 - 3x)/2$, $0 \leqslant x \leqslant x_1 = \frac{1}{6}$ restricts the admissible range of c to $(-\frac{4}{3},\frac{2}{3})$ and for this range, $y(\frac{1}{6}) \leqslant 1$ as can be verified by simple estimates.

On the other hand, for any current function $e(x)$, $0 \leqslant x \leqslant x_1$, the value $y_1 = \int_0^{x_1} e(x)\,dx$ is always attainable with the choice of $c = 0$. Thus our analysis shows that, in this case, the curve whose graph is represented by the function $\tilde{y}(x) = \int_0^x e(t)\,dt$ provides uniquely the least time of transit among all those representable by functions in $\mathfrak{D}$ (or in $\hat{\mathfrak{D}}$, as in Remark 2). Moreover, since the term in braces in (5) vanishes identically when $y' = \tilde{y}' = e$, an obvious modification of the previous argument shows that $\tilde{y}$ also minimizes T uniquely on

$$\tilde{\mathfrak{D}} \overset{\text{def}}{=} \{\tilde{u} \in C^1[0,x_1]: \tilde{u}(0) = 0\}$$

and hence provides (uniquely) the minimum transit time from the origin to the opposite shore regardless of the downstream point of destination. (Physically, this solution corresponds to always steering the boat perpendicular to the river banks.)

Observe also that, in the case of no current or constant current, the boundary conditions can always be met and the minimum transit time is given (uniquely) by the straight line path.

Plateau's Problem. Multidimensional problems also yield to direct attack, as is illustrated by our final example, the important problem of Plateau, which consists of finding a surface of least area spanning a given contour in $\mathbb{R}^3$. We restrict our attention to the nonparametric case of surfaces given by the graph of a smooth function $z = z(x,y)$ defined on D, a bounded planar domain with smooth boundary ∂D, and consider minimizing the area

$$A(z) = \iint_D \left(1 + z_x^2 + z_y^2\right)^{1/2} dx\,dy \quad \text{over} \quad \mathfrak{D} \stackrel{\text{def}}{=} \left\{z \in C^1(\bar{D}) : z|_{\partial D} = g\right\}$$

where $\bar{D} = D \cup \partial D$ and g is a given smooth function on ∂D.

It suffices to find that z which makes

$$\Delta A(z;v) = A(z+v) - A(z) \geqslant 0 \;\forall v \in \mathfrak{D}_0 = \left\{u \in C^1(\bar{D}) : u|_{\partial D} = 0\right\}.$$

Again applying the Cauchy-Schwarz vector inequality as in the proof of the Proposition, we see that for all $v \in \mathfrak{D}_0$

$$\Delta A(z;v) \geqslant \iint_D (Uv_x + Wv_y)\,dx\,dy$$

where

$$U = z_x\left(1 + z_x^2 + z_y^2\right)^{-1/2} \quad \text{and} \quad W = z_y\left(1 + z_x^2 + z_y^2\right)^{-1/2},$$

with equality iff $v \equiv 0$. [Equality $\Rightarrow v_x = v_y = 0$ (on the domain D) $\Rightarrow v = \text{const.} = v|_{\partial D} = 0$.]

Again as in Remark 3 we may seek $z \in \mathfrak{D} \cap C^2(D)$ so that by Green's theorem

$$\Delta A(z;v) \geqslant \int_{\partial D} (Uv\,dx - Wv\,dy) - \iint_D v\left[U_x + W_y\right] dx\,dy.$$

The first integral vanishes for all $v \in \mathfrak{D}_0$ and the second must also when $U_x + W_y = 0$. Any solution to this last equation which satisfies the boundary conditions provides the *unique* minimizing surface (even among corresponding piecewise smooth surfaces). The equation $U_x + W_y = 0$ is equivalent to

$$\left(1 + z_y^2\right)z_{xx} - 2z_x z_y z_{xy} + \left(1 + z_x^2\right)z_{yy} = 0, \tag{6}$$

which is the familiar minimal surface equation obtained from the Euler-Lagrange necessary conditions ([**3**, pp. 599–613]).

Our uniqueness argument shows that this equation cannot have more than one solution $z \in \mathfrak{D} \cap C^2(\bar{D})$ and thus that the Dirichlet problem for this equation has at most one solution. If the domain D is not convex, it is known that the Dirichlet problem for equation (6) is not always solvable. ([**4**, p. 100]) In this case we can draw no additional conclusions from the analysis given here. However, it is also known that if D is convex then the Dirichlet problem for (6) is always solvable, and hence by our argument the associated surface has (uniquely) the minimal area among all those representable by functions z in $\mathfrak{D}$ (or in $\hat{\mathfrak{D}}$, the corresponding class of piecewise C^1 functions).

References

1. G. A. Bliss, Calculus of Variations, Open Court, Chicago, 1925.

2. R. Courant, Calculus of Variations, Institute of Mathematical Sciences, New York University Lecture Notes, New York, 1957.

3. P. R. Garabedian, Partial Differential Equations, Wiley, New York, 1964.

4. R. Osserman, A Survey of Minimal Surfaces, Van Nostrand Reinhold, New York, 1969.

5. D. R. Smith, Variational Methods in Optimization, Prentice-Hall, Englewood Cliffs, N.J., 1974.

6. L. C. Young, Lectures on the Calculus of Variations, W. B. Saunders, Philadelphia, 1969.

MISCELLANEA

54. There is nothing mysterious, as some have tried to maintain, about the *applicability* of mathematics. What we get by abstraction from something can be returned!

—R. L. Wilder, *Introduction to the Foundations of Mathematics*, 1952, p. 275.

CAYLEY'S VERSION OF THE RESULTANT OF TWO POLYNOMIALS

H. B. GRIFFITHS
Faculty of Mathematical Studies, University of Southampton, Southampton, S09 5NH, England

1. Introduction. We consider an algebraic problem which arises from considering pairs of differential equations on the plane of the form

$$\dot{x} = P(x,y), \qquad \dot{y} = Q(x,y)$$

where P and Q are polynomials in x and y. To find the equilibrium points we have to find the intersections of the curves $P(x,y) = 0 = Q(x,y)$, and (among other things) to decide whether they touch there and whether the discriminant $D(x,y)$ vanishes there. (The precise form of D does not matter here: it is again a polynomial.) The first problem reduces to finding when a certain polynomial $f(x)$ in the single variable x has a repeated (real) zero, while the second requires a decision as to whether two polynomials $g(x),h(x)$ have a common root. Algebraically speaking, the second problem includes the first, since $f(x)$ has a repeated zero if and only if $f(x)$ has a root in common with its derivative $f'(x)$.

This paper is written from the point of view of a working mathematician confronted with a problem of which he was initially ignorant; I was led to make inquiries, prove things, hear of known work, axiomatize, modify the earlier proofs, and obtain extensions of them which I believe to be new.

As a first step in finding an exposition of classical nineteenth-century techniques for this algebraic problem (or indeed any other), an invaluable source is always the famed thirteenth edition (1911) of the *Encyclopaedia Britannica*. Its mathematical authors include Cayley, Rayleigh, Maxwell, and Russell, among others, and all were allowed the space to write with considerable technical detail. One therefore looks up the article "Algebraic Forms" by the combinatorialist Major P. A. MacMahon and finds, sure enough, that he discusses the "eliminant" or, equivalently, the *resultant* of two polynomials $f(x)$ and $g(x)$. The resultant is obtained by eliminating x from the equations $f(x) = 0 = g(x)$, and is a multinomial S in the coefficients of f and g. MacMahon gives various ways of calculating S, and each comes from evaluating some determinant; the best known uses Sylvester's method involving a determinant of order $m + n$, where

$$\text{degree}\,(f) = n, \qquad \text{degree}\,(g) = m.$$

However, MacMahon gives a method due to Cayley, involving a determinant Δ of order $\max(m,n)$, which is therefore likely to be less laborious to calculate. His exposition of it is very brief, and not easy to follow because of Victorian neglect of quantifiers. More important, he derives the equation $\Delta = 0$ (like $S = 0$) as a *necessary* condition for f and g to have a common (possibly complex) root, whereas in the problem above one wishes to use it as a *sufficient* condition. The first object of the present article is to give a proof of the sufficiency (see Proposition 1 below).

A satisfactory account of S is given in van der Waerden [**6**, Sections 27, 28] and is valid when the coefficients of f and g lie in an arbitrary field. One could give a similar treatment for Δ or, as R. M. Guralnick has pointed out, use the known sufficiency of S along the following lines. Both S and Δ are multinomials in the coefficients of f and g, and if S vanishes then f and g have a

---

The author received his Ph.D. from Manchester under M. H. A. Newman. After lecturing at Aberdeen, Bristol, and Birmingham, he became a Professor at Southampton in 1964; he has been a visitor at the Institute for Advanced Study, the Courant Institute, and the Institut des Hautes Etudes Scientifiques. His books range from *A Companion to A-level Mathematics* (with J. V. Armitage) and *Mathematics: Society and Curricula* (with A. G. Howson) to *Classical Mathematics* (with P. J. Hilton) and *Surfaces*. His interests are in applications of topology, dynamical systems, and mathematical education. His Centenary Address to the British Mathematical Association appeared in this MONTHLY, 78 (1971), 1067–1085.—*Editors*.

common zero (by the van der Waerden exposition); so by Cayley's equation, Δ vanishes. Hence S must be a factor of Δ, and a suitable comparison of coefficients shows that

$$\Delta = (-1)^{\sigma} f_n^{n-m} S \tag{1.1}$$

where $\sigma = n(n-1)/2$ and f_n is the (nonzero) coefficient of x^n in f. Thus Cayley's condition is also sufficient. However, this argument is essentially that by which van der Waerden relates S to the product

$$G = g(\alpha_1)g(\alpha_2)\cdots g(\alpha_n) \tag{1.2}$$

where the α_i are the n roots of f in the field under consideration, and we count them according to their multiplicities. For G is a symmetric function of the α_i, and hence a multinomial in the coefficients of both f and g which can be compared with the multinomial S.

In the present paper, it is our concern to give a proof for Δ which does not need as much algebraic background, since the latter is often missing from courses of Modern Algebra. (It is not mentioned in the oft-emulated survey of Birkhoff and Mac Lane [**1**], nor is the theory of G included in Bôcher's treatment of S in [**2**], and is thus unlikely to be at the fingertips of a nonspecialist.) The cost of directness is that we must work in the field $\mathbb{C}$ of complex numbers in order to allow the use of continuity arguments. In the algebraic proofs mentioned above, the necessity of Cayley's condition was used in order to prove its sufficiency: the same is true of our proof, but in a different way. Guralnick's comment led me to abstract the features of my original proof in order to clarify its logical structure. The resulting abstraction then led (with some more work) to stronger results about Cayley's *matrix*—its nullity and diagonalizability. To help readability and not cover up motivation, however, we shall prove the weaker result first, in the notation of (1.1) and (1.2) as

PROPOSITION 1. $\Delta = (-1)^{\sigma} f_n^n G$.

From this equation and the definition of G in (1.2), it is clear that $\Delta = 0$ iff some $g(\alpha_i) = 0$ iff f and g have a common root. The proof will be given in Section 6; and since it involves the eigenvalues of a certain matrix, it suggests the possibility of a main theorem (in Section 8) about matrices rather than determinants. One consequence of that work can be stated here without further technical language; if we denote Cayley's matrix by $C(f,g)$, then

$$\text{Nullity}\, C(f,g) = \text{degree}\,(HCF(f,g)) \qquad (g \neq 0)$$

which implies the determinant criterion for f and g to have a common root.

Our working in $\mathbb{C}$ rather than in a general field may also perhaps attract the reader's attention to future work that needs to be done to obtain a *geometrical* understanding of these problems. This is particularly needed when we restrict ourselves to *real* polynomials f,g with coefficients that we assign as coordinates in $\mathbb{R}^{n+1}$. In that case, the set $\Delta = 0$ is a locus L in $\mathbb{R}^{2n+2}$ whose geometry is not well understood: the Victorians were largely content simply to encapsulate its description in the form of a determinant. In the special case that $g = f'$, we may assume that f is of reduced form $f_0 + \cdots + ax^{n-2} + x^n$, with coordinates (and L) therefore in $\mathbb{R}^{n-1}$; and then, when $n \leqslant 6$, the geometry of L has been vigorously pursued recently by workers in catastrophe theory, who have given L the names *fold*, *cusp*, *swallow-tail*, *butterfly*, and *wigwam*, respectively (see, e.g., Poston and Stewart [**5**, p. 74]).

Having thus placed our problem in a more meaningful (and fashionable) setting, we shall clarify MacMahon's exposition in the next section, and in Section 3 we show that Proposition 1 is really symmetrical in f and g. In Sections 4 and 5 we abstract the properties of Cayley's matrix that we need. These properties lead in Section 6 to an abstract version and proof of Proposition 1, which suggests some uniqueness characterizations of determinants and matrices like Cayley's. These characterizations are discussed and proved in Sections 7 and 8, culminating in Theorem 8.1.

2. Cayley's Method. The approach of Cayley, as outlined by MacMahon, is the following. Starting with polynomials f,g with coefficients in $\mathbb{C}$, he forms the polynomial

$$H(x,y)=f(x)g(y)-f(y)g(x) \tag{2.1}$$

from which he factors out $x-y$ to obtain

$$h(x,y)=H(x,y)/(x-y)=\sum a_{rs}x^r y^s \tag{2.2}$$

where the coefficients a_{rs} are in $\mathbb{C}$, and r,s run from 0 to $n-1$; and we suppose that

$$n=\max(\deg f,\deg g)>0. \tag{2.3}$$

Thus (a_{rs}) is Cayley's $(n\times n)$ matrix, symmetric since

$$H(x,y)=-H(y,x). \tag{2.4}$$

Collecting terms, we can also write h as a polynomial in x:

$$h(x,y)=\sum_{r=1}^{n-1}x^r P_r(y)=h_y(x)\text{ (say)} \tag{2.5}$$

where $P_r(y)=\Sigma_s a_{rs}y^s$; each P_r is a polynomial in y of degree $\leqslant n-1$. Now suppose that f and g have a common root z. Then for all $x\in\mathbb{C}$, $H(x,z)=0$; so for all x, except possibly at $x=z$, we have $h(x,z)=0$. Hence the polynomial $h_z(x)$ in (2.5) is the zero polynomial (as a polynomial in x). Therefore

$$P_j(z)=0,\qquad j=0,1,\dots,n-1;$$

so if we now regard these n equations as homogeneous linear equations with nontrivial solution $1,z,\dots,z^{n-1}$, we must have the *necessary* condition on Cayley's determinant:

$$\Delta=\det(a_{rs})=0 \tag{2.6}$$

for f and g to have the common root z. As stated in the Introduction, its sufficiency follows from Proposition 1, the proof of which requires the following preliminaries.

3. Roots of f and g. The equation forming Proposition 1 is in fact symmetrical in f and g, as we show in the following Lemma. As in equation (1.2) above, we enumerate the roots of f and g in $\mathbb{C}$ (repeating multiple roots) as, respectively,

$$\alpha_1,\alpha_2,\dots,\alpha_n,\qquad \beta_1,\beta_2,\dots,\beta_m, \tag{3.1}$$

where we assume that $n\geqslant m>0$. Thus if the leading coefficients of f and g are f_n,g_m then

$$f(x)=f_n\cdot(x-\alpha_1)\cdots(x-\alpha_n),\qquad g(x)=g_m\cdot(x-\beta_1)\cdots(x-\beta_m). \tag{3.2}$$

Let

$$F=f(\beta_1)f(\beta_2)\cdots f(\beta_m),\qquad G=g(\alpha_1)g(\alpha_2)\cdots g(\alpha_n).$$

3.3. LEMMA.

$$G=g_m^n\prod_{i,j}(\alpha_i-\beta_j)=(-1)^{nm}F.$$

Proof. The equations (3.2) show that $g(\alpha_1)$, say, is $g_m\cdot(\alpha_1-\beta_1)\cdots(\alpha_1-\beta_m)$. Hence, if each factor of G is thus factorized we find $G=g_m^n\Pi_{i,j}(\alpha_i-\beta_j)$, and analogously for F but with $\beta_j-\alpha_i$ replacing $\alpha_i-\beta_j$. Hence we obtain the first equation above; for the second there are nm brackets giving the power $(-1)^{nm}$ as required. ■

In trying to prove the sufficiency of Cayley's condition, an obvious procedure is to assume that f and g do *not* have a common root, even though $\Delta=0$, and then see what happens if we perturb

g to $g-\lambda$, where $\lambda \in \mathbf{C}$. This leads to the basic observation that $g-\lambda$ and f have a common root iff $\lambda = g(\alpha)$ for some root α of f. Hence by iteration we obtain the following Lemma, of which the Corollary will not be needed until Section 8.

3.4. LEMMA. *For each $\lambda \in \mathbf{C}$, f and $g-\lambda$ have a (monic) highest common factor of the form*

$$h = (x-\alpha_1)^{p(1)}(x-\alpha_2)^{p(2)}\cdots(x-\alpha_k)^{p(k)}$$

where the α's are distinct roots of f such that

$$\{\alpha_1,\alpha_2,\ldots,\alpha_k\} \subseteq g^{-1}(\lambda).$$

(We take $h=1$ if $k=0$). ■

Define the integer $\mu(\lambda)$ to be the degree of h, so

$$\mu(\lambda) = p(1)+p(2)+\cdots+p(k), \tag{3.5}$$

where $\mu(\lambda)=0$ if $k=0$. (We need not incorporate f and g into the notation since they will be fixed throughout the discussions.)

Note that at most n numbers λ can be nonzeros of μ because each such λ must be of the form $g(\alpha_i)$ for some root α_i of f. Further, $p(i)$ cannot exceed the multiplicity of α_i (as a root of f), so

$$n \geqslant \sum_{\lambda} \mu(\lambda), \tag{3.6}$$

the only essential contributors to the sum being the nonzeros of μ.

We shall say that g *matches* f provided we have equality in (3.6); clearly, in that case, each $p(i)$ in (3.5) equals the multiplicity of α_i as a root of f. In particular, then, we have:

3.7. COROLLARY. *If the roots of f are all distinct, then for all $\lambda \in \mathbf{C}$, $\mu(\lambda)=0$ or 1, and g matches f.* ■

4. The Cayley Matrix. Returning to h in (2.2), let us now indicate its dependence on f and g by writing h as a "product": $h = f*g$. From the definitions it is clear that this product is anti-commutative and bilinear, that $f*f=0$ and for any constant c

$$f*(cg) = c(f*g).$$

If further we write $M(f*g)$ for the Cayley matrix (a_{rs}) in equation (2.2), then

$$\det M(f*g) = \det(a_{rs}) \tag{4.1}$$

and

$$M(f*g) = -M(g*f).$$

In particular, let $\mathbf{1}$ denote the constant unit polynomial ($g(x)\equiv 1$); then if $f(x) = f_0 + f_1x + \cdots + f_nx^n$ with $f_n \neq 0$, it is easy to verify that

$$M(f*\mathbf{1}) = \begin{vmatrix} f_1 & f_2 & \cdots & f_{n-1} & f_n \\ f_2 & f_3 & \cdots & f_n & 0 \\ \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ f_{n-1} & f_n & 0 & \cdot & \cdot \\ f_n & 0 & 0 & \cdot & 0 \end{vmatrix}. \tag{4.2}$$

Hence, writing $\sigma = n(n-1)/2$, we obtain by induction on n that

$$\det M(f*\mathbf{1}) = (-1)^{\sigma} f_n^n \neq 0. \tag{4.3}$$

As a notational point, we shall find it convenient to think of $(f_0, f_1, \ldots, f_n)$ as the coordinates of f, so that we may use $\mathbf{x}^r$ to mean the r-plus-first basis vector in $\mathbb{C}^n$. Then f is the vector sum $f = f_0 + f_1\mathbf{x} + \cdots + f_n\mathbf{x}^n$ and "$f(x)$" is not needed.

Direct computation shows that $M(\mathbf{x}^n * \mathbf{x}^m)$ is a matrix of zeros and ones; so the bilinearity of the product implies that each entry in $M(f*g)$ is a quadratic form (without square terms) in the coefficients of f and g. Hence $\det M(f*g)$ is a multinomial in these coefficients, of degree $\leqslant n$ in any single one.

In Section 8 below we shall need the following result. For each $t \in \mathbb{C}$, let $T_t f$ denote the "translate" of f, given by

$$(T_t f)x = f(x+t)$$

expressed as a new polynomial in x. The formula for h in (2.2) above shows at once that

$$T_t f * T_t g = T_t(f*g) \tag{4.4}$$

where we write $T_t h$ for the translate given by

$$(T_t h)(x,y) = h(x+t, y+t).$$

5. The Cayley Mapping $\Gamma: \mathcal{P}_n \to \mathcal{M}_n$. The structure of the proof of Proposition 1 will be clearer if we first abstract here those properties of the Cayley matrix $M(f*g)$ that are actually needed. For this purpose let $\mathcal{P}_n, \mathcal{M}_n$ denote the sets of complex polynomials of degrees $\leqslant n$, and $n \times n$ matrices, respectively. Then if we fix f of degree n, we obtain a map $\Gamma: \mathcal{P}_n \to \mathcal{M}_n$ given by

$$\Gamma(g) = M(f*g).$$

We call Γ the *Cayley mapping* (relative to f).

After (4.3) we mentioned that we would regard the coefficients of a polynomial in $\mathcal{P}_n$ as its coordinates in $\mathbb{C}^{n+1}$. Similarly, we use a coordinate chart in $\mathcal{M}_n$ by regarding the entries of a matrix as its coordinates. It then makes sense to assert that

$$\Gamma: \mathcal{P}_n \to \mathcal{M}_n$$

is clearly *continuous*.

As we have already seen, Γ satisfies the following conditions, by (4.2), by linearity of the Cayley product $f*g$, and by (2.6), respectively:

$\mathbf{C}_1$. *$\Gamma(\mathbf{1})$ is an invertible matrix.*

$\mathbf{C}_2$. *For each $g \in \mathcal{P}_n$ and $c \in \mathbb{C}$*

$$\Gamma(g+c) = \Gamma(g) + c\Gamma(\mathbf{1}).$$

$\mathbf{C}_3$. *If f and g have a common zero in $\mathbb{C}$, then $\Gamma(g)$ is not invertible.*

Any map $\phi: \mathcal{P}_n \to \mathcal{M}_n$ satisfying conditions $\mathbf{C}_1$-$\mathbf{C}_3$ will be called an *f-map*, and we may as well prove a version of Proposition 1 for abstract f-maps (which turn out not to be so abstract after all).

6. f-maps. Keeping f fixed as in conditions $\mathbf{C}_1$-$\mathbf{C}_3$, consider some properties of f-maps in general. For example, the Cayley mapping Γ is an f-map; and if U, V are arbitrary invertible matrices in $\mathcal{M}_n$, then from one f-map ϕ we obtain another, given by

$$g \to U \cdot \phi(g) \cdot V,$$

using matrix multiplication in $\mathcal{M}_n$.

In any case, if we pass to determinants, we obtain a type of uniqueness theorem as follows. Let

$$\psi^+ = \det \psi(\mathbf{1});$$

recall that Γ^+ was evaluated in (4.3). Then we shall prove the following proposition, which shows

that $\det\psi(g)$ depends only on ψ^+ and G in (1.2).

PROPOSITION 2. *Suppose that all the n roots of f are distinct. Then any continuous f-map* $\psi:\mathcal{P}_n\to\mathfrak{M}_n$ *satisfies*

$$\det\psi(g)=\psi^+\cdot G.$$

This proposition has two important Corollaries, one of which is Proposition 1, which we stated in the Introduction.

COROLLARY 1. ψ *satisfies the converse of condition* $\mathbf{C}_3$.

COROLLARY 2. *Proposition* 2 *implies that for any polynomials* f,g *Proposition* 1 *holds.*

Let us first deduce the corollaries from the main assertion. The first follows at once by the remarks following the enunciation of Proposition 1 in the Introduction. For the second corollary, we need consider only the case when f has a multiple root. Let $f_k=f+1/k$, $k=1,2,\dots$. Now z would be a multiple root of f_k, provided both $f(z)=-1/k$ and $f_k'(z)=0$; but $f_k'=f'$, and at most $n-1$ integers k can satisfy $f(z)=-1/k$ if z is a root of f'. Hence there exists K such that, for all $k>K$, f_k has distinct roots. Let Γ_k denote the Cayley map associated with f_k, and let $F_k=f_k(\beta_1)f_k(\beta_2)\cdots f_k(\beta_m)$, where the β's are the roots of g. Then by Lemma 3.1, Proposition 2 implies (when $k>K$)

$$\det\Gamma_k(g)=(-1)^{nm}\Gamma_k^+\,F_k. \tag{6.1}$$

By continuity of all three functions det, M, and $f*g$, we may take limits as k tends to infinity and obtain

$$\det\Gamma(g)=(-1)^{nm}\Gamma^+F=\Gamma^+G \tag{6.2}$$

which is Proposition 1 since $\Gamma(g)=M(f*g)$ with determinant Δ, and $\Gamma^+=(-1)^\sigma f_n^n$ by (4.2). ■

The way in which we deduce (6.2) from (6.1) suggests how we may strengthen the hypotheses on ψ in Proposition 2 so as to eliminate the condition that f has distinct roots. We must then add an axiom that allows the existence and convergence of f_k-maps ψ_k analogous to the maps Γ_k in (5.1) above. The most obvious axiom is suggested by the Cayley matrix $M(f*g)$: we require a "product" map $\theta:\mathcal{P}_n\times\mathcal{P}_n\to\mathfrak{M}_n$ such that for each $f\in\mathcal{P}_n$ the map

$$g\to\theta(f,g),\qquad \mathcal{P}_n\to\mathfrak{M}_n$$

is an f-map. Then the previous mode of deduction of equation (6.2) gives at once

PROPOSITION 3. *If* θ *is continuous, then for all* $(f,g)\in\mathcal{P}_n\times\mathcal{P}_n$,

$$\det\theta(f,g)=G\cdot\det\theta(f,\mathbf{1}). \qquad ■$$

Let us now prove Proposition 2 itself.

Proof of Proposition 2. By condition $\mathbf{C}_1$, $\psi(\mathbf{1})^{-1}$ exists. Thus we observe that *each number* $g(\alpha_i)$ *is an eigenvalue of the matrix* $W=\psi(\mathbf{1})^{-1}\psi(g)$. For the polynomial $g-g(\alpha_i)$ has the zero α_i in common with f, so if I_n denotes the unit matrix in $\mathfrak{M}_n$, we have

$$\begin{aligned}0&=\det(\psi(g-g(\alpha_i)))\qquad\text{by condition }\mathbf{C}_3\\&=\det\left[\psi(\mathbf{1})(W-g(\alpha_i)I_n)\right]\qquad\text{by }\mathbf{C}_2\end{aligned}$$

so

$$0=\det(W-g(\alpha_i)I_n)$$

by the multiplicative property of det, since $\det\psi(\mathbf{1})\neq0$ by $\mathbf{C}_1$. Hence $g(\alpha_i)$ is an eigenvalue of W.

Now suppose that the numbers $g(\alpha_1),\dots,g(\alpha_n)$ are *all distinct*. Since there are n of them and $W \in \mathfrak{M}_n$, then they constitute *all* the eigenvalues of W; so

$$\det W = g(\alpha_1)g(\alpha_2)\cdots g(\alpha_n).$$

But

$$\det W = \det\psi(g)/\det\psi(\mathbf{1})$$

so

$$\det\psi(g) = \psi^+\cdot g(\alpha_1)\cdots g(\alpha_n) \tag{6.3}$$

Finally, suppose that the numbers $g(\alpha_i)$ are *not* all distinct. Let $g_k(x) = g(x) + x/k$, $k = 1,2,\dots$. Then since the numbers α_i themselves are all distinct, none of the numbers $(g(\alpha_i) - g(\alpha_j))/(\alpha_j - \alpha_i)$ can be of the form $1/k$ if k is sufficiently large, say if $k > k_0$. Therefore, for each $k > k_0$, the numbers $g_k(\alpha_i)$ are distinct, so by (6.3) above

$$\det\psi(g_k) = \psi^+\cdot g_k(\alpha_1)\cdots g_k(\alpha_n). \tag{6.4}$$

Hence, since ψ is continuous, we may take limits as k tends to infinity, and obtain (6.3) for all $g \in \mathcal{P}_n$. This completes the proof of Proposition 2. ∎

The last proof leads to the following observations.

REMARK 1. It was vital, in the derivation of equation (6.3), to have $\psi(g) \in \mathfrak{M}_n$, otherwise we would not have known all the eigenvalues of W. We do not know how to circumvent this point, to characterize f-maps $\mathcal{P}_n \to \mathfrak{M}_m$ for other integers $m \neq n$.

REMARK 2. It is interesting that the linear nature of the product $f*g$ in (4.1) needed to be reflected only in the apparently weaker condition $\mathbf{C}_2$. As to the anti-commutativity in (4.1), the way in which it links values of $\det M(f_n*\mathbf{1})$ and $\det M(\mathbf{x}^n*\mathbf{1})$ is brought out in the following "abstract" calculation. Suppose that there is an $\mathbf{x}^n$-map $\varepsilon:\mathcal{P}_n \to \mathfrak{M}_n$ which satisfies the equation of Proposition 2 (even though $\mathbf{x}^n$ does not have distinct roots: but such a map could be $g \to \theta(\mathbf{x}^n,g)$, by Proposition 3). Then by that equation

$$\det\varepsilon(f) = \varepsilon^+\cdot f(0)^n$$

whereas any continuous f-map ψ satisfies

$$\det\psi(\mathbf{x}^n) = \psi^+\cdot\alpha_1^n\alpha_2^n\cdots\alpha_n^n = \psi^+\cdot(-1)^n(f(0)/f_n)^n.$$

Hence

$$\frac{\det\varepsilon(f)}{\varepsilon^+} = (-f_n)^n\frac{\det\psi(\mathbf{x}^n)}{\psi^+}, \tag{6.5}$$

an equation devoid of content when $f(0) = 0$, for then each side is zero by condition $\mathbf{C}_2$. However, when ε and ψ both arise from the Cayley product, we have

$$\varepsilon(f) = M(\mathbf{x}^n*f) = -M(f*\mathbf{x}^n) = -\psi(\mathbf{x}^n),$$

so the anti-commutativity implies here that $\psi^+ = f_n^n\varepsilon^+$, agreeing with the calculation given in (4.2). But in the general case, equation (6.5) shows that for any continuous f-map ψ, the composition $\mathcal{P}_n \xrightarrow{\psi} \mathfrak{M}_n \xrightarrow{\det} \mathbf{C}$ is essentially determined by its effect on either of the two "extreme" polynomials $\mathbf{1},\mathbf{x}^n$—for then we know ψ^+ in (6.3).

7. Diagonalization and $\psi(g)$. The deduction of equation (6.3) would allow us to assert more than we did: we could have asserted that W is similar to the diagonal matrix

$$D_g = \operatorname{diag}(g(\alpha_1),g(\alpha_2),\dots,g(\alpha_n)) \tag{7.1}$$

whose entries are all zero except for the n distinct numbers $g(\alpha_1),\dots,g(\alpha_n)$ down the principal

diagonal. Thus W is of the form VD_gV^{-1} for some $V \in \mathfrak{M}_n$, and W is *diagonalizable*. The question arises whether W is diagonalizable for all g; and we cannot use the limiting argument we gave for equation (6.4) unless we know in advance that V always lies in some compact subset of $\mathfrak{M}_n$.

We can in any case say that W is diagonalizable for "almost all" g in $\mathcal{P}_n$, in the following sense. The possibly exceptional polynomials g will be those for which there exist distinct roots α_i, α_j of f such that $g(\alpha_i) = g(\alpha_j)$. In $\mathcal{P}_n$, this set L_{ij} is a linear subspace, and the exceptional set is therefore included in the union L of the L_{ij} over all possible i,j. Thus L has (topological) dimension 2 less than that of $\mathcal{P}_n$ and is nowhere dense in $\mathcal{P}_n$.

If we are dissatisfied with such a "generic" conclusion, we can instead strengthen our hypotheses to guarantee that W is always diagonalizable (perhaps restricting f). This would then give a type of uniqueness theorem at the matrix level since $W = \psi(\mathbf{1})^{-1}\psi(g)$, by saying that for *all* g, $\psi(g)$ would have the form $\psi(\mathbf{1})\cdot VD_gV^{-1}$ with D_g as in (7.1). We must therefore look more closely at the case of multiple eigenvalues of W.

Now, when W has repeated eigenvalues λ of multiplicity $m(\lambda)$ as in (5.4), a standard theorem (see, e.g., Mirsky [**4**, Th. 10.2.3, p. 294]) tells us that W is still diagonalizable iff the following condition holds: for each eigenvalue λ, we require the corresponding eigenspace E_λ to have dimension $m(\lambda)$ over $\mathbb{C}$. But

$$\begin{aligned} E_\lambda &= \{v \in \mathbb{C}^n \mid Wv = \lambda v\} \\ &= \{v \in \mathbb{C}^n \mid \psi(\mathbf{1})\cdot(W - \lambda I_n)c = 0\} \end{aligned}$$

since $\psi(\mathbf{1})$ is invertible. Hence E_λ is here exactly the kernel of $\psi(g-\lambda)$, by condition $\mathbf{C}_2$, since $W = \psi(\mathbf{1})^{-1}\psi(g)$. Therefore

$$\dim E_\lambda = \dim(\ker(\psi(g-\lambda)).$$

It is customary to define the *nullity* of a matrix $A \in \mathfrak{M}_n$ to be dim(ker(A)), and we denote it by $N(A)$. By Sylvester's "law of nullity,"

$$N(A) = n - \operatorname{rank}(A), \tag{7.2}$$

and hence we always have (see Mirsky [**4**, p. 214]

$$m(\lambda) \geq N(W - \lambda I_n).$$

Thus, by the calculation above

$$m(\lambda) \geq \dim E_\lambda = N(\psi(g-\lambda)). \tag{7.3}$$

Now the sum of the multiplicities $m(\lambda)$ is n; so

$$n \geq \sum_\lambda N(\psi(g-\lambda)), \tag{7.4}$$

the last sum being taken over all distinct eigenvalues λ of W. Suppose therefore that we could prove that (7.4) is in fact an equality: then for each λ in (7.3) we would also have equality, and W would be diagonalizable. Conversely, if W is diagonalizable we have equality in (7.3) and hence in (7.4).

Recall from (3.5) that g was said to *match* f if we have equality there: $n = \Sigma_\lambda \mu(\lambda)$. Hence we would also have the equality we want in (7.4) if g were matched to f and we could prove that the f-map ψ satisfies the

7.5. NULLITY CONDITION. *For each* $\lambda \in \mathbb{C}$, $N(\psi(g-\lambda)) = \mu(\lambda)$.

To summarize, if we can find suitable extra conditions on ψ for the nullity condition to be valid, then we shall have equality in (7.4) whenever g matches f. In particular, that will be for *all* $g \in \mathcal{P}_n$ whenever f has n distinct roots, by Corollary 3.6. The extra conditions ought, of course, to

be verifiable when ψ is the Cayley map Γ. In the next section we set up the conditions, which in view of 7.5 it is natural to express in terms of nullity.

8. Product Mappings $\mathcal{P}_n \times \mathcal{P}_n \to \mathfrak{M}_n$. As with Proposition 3, we shall suppose that we have a product map

$$\vee : \mathcal{P}_n \times \mathcal{P}_n \to \mathfrak{M}_n, \qquad (f,g) \to f \vee g$$

satisfying the following conditions, in which T_t denotes the translation operator that was used in (4.4); we require:

$\mathbf{P}_1$. $N((\mathbf{x}f) \vee (\mathbf{x}g)) = N(f \vee g) + 1$ (*if* $f,g \in \mathcal{P}_{n-1} \subseteq \mathcal{P}_n$).
$\mathbf{P}_2$. *For all* $t \in \mathbf{C}$, $N(T_t f \vee T_t g) = N(f \vee g)$ (*if* $f,g \in \mathcal{P}_n$).
$\mathbf{P}_3$. *The map* $\psi_f: \mathcal{P}_n \to \mathfrak{M}_n$, *given by* $g \to f \vee g$, *is an f-map, and satisfies the converse of* $\mathbf{C}_3$.

We shall call such a product map a *nullity product.* These rather weird-looking conditions are in fact satisfied by the Cayley product $M(f*g)$ as we show in Lemma 8.9 below. Their usefulness lies in our being able now to formulate our main theorem.

8.1. THEOREM. *Let* $\psi_f: \mathcal{P}_n \to \mathfrak{M}_n$ *be derived from a nullity product. Then for all* $g \in \mathfrak{M}_n$, g *matches f iff there exists* $Q_g \in \mathfrak{M}_n$ *such that*

$$\psi_f(g) = \psi_f(\mathbf{1}) \cdot Q_g D_g Q_g^{-1}$$

where $D_g = \operatorname{diag}(g(\alpha_1), g(\alpha_2), \ldots, g(\alpha_n))$ *and the* α*'s are the roots of f counted with their multiplicities.*

Since g matches f when the roots of f are all distinct (by Corollary 3.6), then Proposition 2 for ψ_f follows from this theorem if we take determinants. The proof of the theorem needs two lemmas, the first of which was quoted in the Introduction, for the special case when $f \vee g$ is taken to be the Cayley product $M(f*g)$. It is presumably related to a strange result of Bôcher [**2**, p. 197].

8.2. LEMMA. *Suppose that the highest common factor of f and g is a (nonzero) polynomial of degree* $r \geqslant 0$. *Then*

$$N(f \vee g) = r.$$

Proof. If f and g have no common root, then $r = 0$; and by condition $\mathbf{P}_3$ above, $\psi(g)$ is invertible with nullity zero.

If f and g have a common root t, then we can write $f = (x-t)f_1$, $g = (x-t)g_1$; so

$$\begin{aligned} N(f \vee g) &= N(T_t f \vee T_t g) \qquad \text{by } \mathbf{P}_2 \\ &= N(\mathbf{x}T_t f_1 \vee \mathbf{x}T_t g_1) \end{aligned}$$

since T_t preserves polynomial multiplication,

$$\begin{aligned} &= N(T_t f_1 \vee T_t g_1) + 1 \qquad \text{by } \mathbf{P}_1 \\ &= N(f_1 \vee g_1) + 1 \qquad \text{by } \mathbf{P}_2. \end{aligned}$$

By extracting successive linear common factors in this way, we obtain

$$N(f \vee g) = N(f_r \vee g_r) + r$$

with f_r and g_r relatively prime. Thus f_r, g_r have no common root, and so $N(f_r \vee g_r) = 0$ by the result when $r = 0$; and the required equation follows. ■

Recall now our need to establish the Nullity Condition (7.5) for ψ_f.

8.3. LEMMA. *If ψ_f is derived from a nullity product, then it satisfies the Nullity Condition* (7.5).

Proof. To prove that $N\psi_f(g-\lambda)=\mu(\lambda)$ we recall the notation of Lemma 3.4 and equation (3.5). Since h there is nonzero, its degree is always $\mu(\lambda)$; so

$$N\psi_f(g-\lambda)=N(f\vee(g-\lambda))=\mu(\lambda)$$

by Lemma 8.2 above. ■

Proof of Theorem 8.1. As we remarked after (7.5), if ψ_f satisfies the Nullity Condition, then $W=\psi_f(\mathbf{1})^{-1}\psi_f(g)$ is diagonalizable iff g matches f. But the proof of Proposition 2 showed that the eigenvalues of W are $g(\alpha_1),g(\alpha_2),\dots,g(\alpha_n)$. This completes the proof of Theorem 8.1. ■

The notation in Theorem 8.1 raises the question: When does Q_g depend continuously on g? We do not know an answer.

Finally, let us show that the Cayley product $M(f*g)$ satisfies conditions $\mathbf{P}_1,\mathbf{P}_2,\mathbf{P}_3$. The first two arise from general properties of polynomials of the form

$$p(x,y)=\sum_{r,s=0}^{n-1}p_{rs}x^ry^s \qquad (n\geq 1)$$

with which we naturally associate the matrix $(p_{rs})=M(p)\in\mathfrak{M}_n$. It is easily verified that $M(\mathbf{xy}\cdot p)$ in $\mathfrak{M}_{n+s}$ is of the form

$$M(\mathbf{xy}\cdot p)=\begin{pmatrix}0&0\\0&M(p)\end{pmatrix}$$

with the same rank as $M(p)$; so by (7.2) the nullity satisfies

$$NM(\mathbf{xy}\cdot p)=NM(p)+1. \tag{8.4}$$

Also, if $t\in\mathbb{C}$ and T_t denotes the translation operator, then

$$(T_tp)(x,y)=p(x+t,y+t)=\sum_{r,s=0}^{n-1}q_{rs}x_ry^s,$$

and it can be verified that

$$(q_{rs})=M(T_tp)=B_t'\cdot M(p)\cdot B_t \tag{8.5}$$

where B_t is a matrix that depends only on n and t; its entries are polynomials in t which involve binomial coefficients. We can avoid the tricky task of evaluating $\det(B_t)$ by the following argument.

8.6. LEMMA. *For all $t\in\mathbb{C}$, B_t is invertible.*

Proof. Fix $t\in\mathbb{C}$ and take p to be the Cayley polynomial $f*g$ of (4.2). By Proposition 1, $M(f*g)$ is invertible iff f and g have a common root, i.e., iff T_tf,T_tg have a common root, i.e., iff $M(T_tf*T_tg)$ is invertible. But by (4.4)

$$M(T_tf*T_tg)=M(T_t(f*g))=B_t'\cdot M(f*g)\cdot B_t \tag{8.7}$$

using (8.5) above. In particular choose $f=\mathbf{x}^n$, $g=\mathbf{1}$, so that f and g have no common root. Therefore the first and third matrices M are each invertible; hence so is B_t. ■

From equation (8.5) we have at once

8.8. COROLLARY. *For all $t\in\mathbb{C}$, $NM(T_tp)=NM(p)$.* ■

We can now prove our final

8.9. LEMMA. *The Cayley product $M(f*g)$ is a nullity product.*

Proof. As in the proof of Lemma 8.6, we take the polynomial $p(x,y)$ to be $f*g$. Then equation (8.7) shows that condition $\mathbf{P}_2$ holds if $f \vee g$ means $M(f*g)$, since B_t is invertible. By definition of $h = f*g$ in (2.2) we compute that

$$(\mathbf{x}f)*(\mathbf{x}g) = \mathbf{x}\mathbf{y}\cdot(f*g)$$

(reading $\mathbf{x}$ and $\mathbf{y}$ as polynomials in two variables). Hence, by equation (8.4), condition $\mathbf{P}_1$ holds for $M(f*g)$. Finally, $\mathbf{P}_3$ holds by Proposition 1, since the Cayley map Γ is an f-map. This proves the lemma. ∎

In view of Corollary 3.7 we can specialize Theorem 8.1 to the "concrete" conclusion:

If f has distinct roots $\alpha_1,\alpha_2,\ldots,\alpha_n$, then for all $g \in \mathcal{P}_n$, $M(f*\mathbf{1})^{-1}\cdot M(f*g)$ is similar to $\operatorname{diag}(g(\alpha_1),\ldots,g(\alpha_n))$. ∎

After the first draft of this paper was complete, Mr. Tony Crilly showed the author the manuscript of his historical essay on Cayley's original treatment [3] of the determinant (a_{rs}) in (2.5). Cayley used a strange notation, and was very brief; he discussed the article later in letters with Sylvester and in other articles. I am grateful to Mr. Crilly; it is clear from his work that the sufficiency of Cayley's condition was not discussed explicitly by the nineteenth-century authors.

References

1. G. Birkhoff and S. Mac Lane, A Survey of Modern Algebra, Macmillan, London, 1965.
2. S. Bôcher, Introduction to Higher Algebra, Macmillan, New York, 1938.
3. A. Cayley, Note sur la méthode d'élimination de Bézout, J. Reine Angew. Math. (Crelle), 53 (1857) 366–367.
4. L. Mirsky, An Introduction to Linear Algebra, Oxford, 1955.
5. T. Poston and I. N. Stewart, Taylor Expansions and Catastrophes, Pitman, 1976.
6. B. L. van der Waerden, Modern Algebra, vol. 1, Ungar, New York, 1953.

CORRECTIONS TO "CYCLOTOMIC POLYNOMIALS AND FACTORIZATION THEOREMS"

[this MONTHLY, 85 (1978) 734–737]

SOLOMON W. GOLOMB
Department of Mathematics, University of Southern California, Los Angeles, CA 90007

In "Cyclotomic Polynomials and Factorization Theorems," the following appears immediately following the proof of Theorem 3:

> For example, $x^{2r} + x^r + 1$ is irreducible over the rationals if and only if $r = 3^k$. Now, all the polynomials $x^{2\cdot 3^k} + x^{3^k} + 1$ remain irreducible modulo 2, as well as modulo any prime q with $q \equiv -1 \pmod 6$, while factoring modulo 3 and modulo q for every prime q with $q \equiv 1 \pmod 6$.

The second sentence is incorrect as printed. The correct statement is:

> Now, all the polynomials $x^{2\cdot 3^k} + x^{3^k} + 1$ remain irreducible modulo q if $q \equiv 2 \pmod 9$ or $q \equiv 5 \pmod 9$, while factoring modulo q for all other primes q.

(The relevant modulus is 9 because $x^6 + x^3 + 1 = \Phi_9(x)$; and the primitive roots modulo 9 are 2 and 5.) I am indebted to Hugh M. Edgar for pointing out the need for this correction, along with the specific example

$$x^6 + x^3 + 1 = (x^2 + 3x + 1)(x^2 + 4x + 1)(x^2 - 7x + 1) \pmod{17},$$

which shows that the original statement is incorrect, since $17 \equiv -1 \pmod 6$.

The reason that $\Phi_{3^{k+1}}(x) = x^{2 \cdot 3^k} + x^{3^k} + 1$ is reducible or irreducible modulo q for the same sets of primes q as $\Phi_9(x) = x^6 + x^3 + 1$, for all $k \geqslant 1$, is that the primitive roots modulo 3^{k+1} are precisely the numbers $g \equiv 2 \pmod 9$ and $g \equiv 5 \pmod 9$. This is a special case of the result [**1**] that, if g is primitive modulo p^2, then g is primitive modulo p^k for all $k \geqslant 2$, for all primes $p > 2$.

Professor John Brillhart of the University of Arizona has pointed out an error in Table 2 on page 737. For $r = 192$, the smallest prime factor should be listed as 17,047,297. According to Brillhart, the incorrect value of 145,143,857, which was furnished by the anonymous referee, occurs as an entry in the manuscript of the *Cunningham Project* (an extensive compilation of number theory tables to be published by the AMS) one line above the correct value for $r = 192$. Also, for $r = 288$, left open in Table 2, the *Cunningham Project* records the prime factor 1,718,990,209.

Reference

1. E. Landau, Vorlesungen über Zahlentheorie, S. Hirzel, Leipzig, 1927, vol. 1, proof of Satz 124, p. 81.

MISCELLANEA

55.

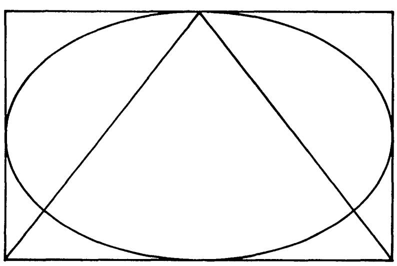
HOMMAGE A ARCHIMEDE

LESTER H. LANGE
School of Science, San José State University, San José, CA 95192

To Robert D. Clark for his seventieth birthday

In the central quad at San José State University, directly across from the landmark tower, there now stands a noteworthy seven foot bronze abstract sculpture, *Hommage à Archimède*, which provides an already pleasant place with an additional pleasant intellectual sweep. Made possible by contributions from friends of the School of Science at the University, the sculpture incorporates several artistic design ideas, a principal one being somewhat reminiscent of the gravestone of Archimedes. Built into the sculpture is a large rectangular figure of "divine proportions," reflecting the observation that if we rotate this figure about its central vertical axis, we find that the resulting volumes satisfy

$$\text{Cone} : \text{Ellipsoid} : \text{Cylinder} = 1:2:3.$$



MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

FAT, SYMMETRIC, IRRATIONAL CANTOR SETS

DUANE BOES
Department of Statistics, Colorado State University, Fort Collins, CO 80523

RICHARD DARST
Department of Mathematics, Colorado State University, Fort Collins, CO 80523

PAUL ERDŐS
Hungarian Academy of Sciences and University of Colorado

A familiar class of symmetric Cantor sets is obtained as follows: Let $\alpha \in (0,1]$; from $[0,1]$ remove a segment of length $\alpha/3$ to leave two intervals of equal length; from each of these intervals remove a segment of length $\alpha/3^2$ to leave 2^2 intervals of equal length; iterate this process and denote the Cantor set that remains by C_α. This class of Cantor sets is a very fruitful source of examples. An introduction to these Cantor sets and their corresponding Cantor functions can be found in [**1**]; however, this note does not depend on [**1**]. This note shows that, except for a first category set of α's in $[0,1]$, $(0,1) \cap C_\alpha$ contains only irrational numbers. Actually, we show that if $x \in (0,1)$ then the set $[x]$ of α's in $[0,1]$ for which $x \in C_\alpha$ is a closed, nowhere dense subset of $[0,1]$; consequently, $\cup_{x\in A}[x]$ is a first category subset of $[0,1]$ whenever A is a countable subset of $(0,1)$. Letting A be the set of rationals in $(0,1)$ produces irrational Cantor sets.

Focus on the construction of C_α and observe that a point $x \in C_\alpha$ is the intersection of a nested sequence of intervals. Thus x is determined (uniquely) by specifying whether a left or a right subinterval contains x at each step: for $0 < \alpha \leqslant 1$ there is a one-to-one correspondence between the elements $x \in C_\alpha$ and the elements $S \in \mathbb{S}$, where $\mathbb{S}$ denotes the set of subsets of the set N of positive integers; $x \in C_\alpha$ corresponds to the set S_x of positive integers n such that x is in a right subinterval at step n. For $0 < \alpha \leqslant 1$, let ϕ_α denote the map that takes $S_x \in \mathbb{S}$ to $x \in C_\alpha$. For future reference, notice that if x_n denotes the left endpoint of the nth step interval that contains x then $x_1 \leqslant x_2 \leqslant \cdots \to x$. For $\alpha = 0$, there may be two subsets of N corresponding to $x \in C_0 = [0,1]$; for example, $\frac{1}{2}$ corresponds to both the one-element set $\{1\}$ and its complement. Nevertheless, we can define $\phi_0 : \mathbb{S} \to C_0$ as we did for $0 < \alpha \leqslant 1$.

For $S \subset N$, let $\lambda(S) = 2\Sigma_{n\in S}3^{-n}$ and $\mu(S) = \Sigma_{n\in S}2^{-n}$; in particular, $\lambda(\phi) = \mu(\phi) = 0$. Then $\lambda(\mathbb{S}) = C_1$ and $\mu(\mathbb{S}) = C_0$; these are the extreme cases $\alpha = 1$ and $\alpha = 0$. Notice that $(\frac{2}{3}-\frac{1}{2}) = \Sigma_{1<n<\infty}(2^{-n} - 2\cdot 3^{-n})$; so $\lambda(S) < \mu(S)$ if $1 \notin S \neq \varnothing$ and $\mu(S) < \lambda(S)$ if $1 \in S \neq N$.

Continue to focus on the construction of C_α. After step one, two intervals of length $l_1 = 2^{-1}(1 - \alpha/3)$ remain; after step two, four intervals of length $l_2 = 2^{-1}(l_1 - \alpha/3^2)$ remain. Continuing, one sees that, after each step n, 2^n intervals of length l_n remain, where

$$\begin{aligned} l_n &= 2^{-1}(l_{n-1} - \alpha/3^n) \\ &= 2^{-n}\big(1 - (\alpha/3)\big[1 + (2/3) + \cdots + (2/3)^{n-1}\big]\big) \\ &= 2^{-n}(1-\alpha) + 3^{-n}(\alpha). \end{aligned}$$

Next notice that if an integer $n \in S \in \mathbb{S}$ and if $x \in C_\alpha$ corresponds to S then $x_{n+1} - x_n = l_n + \alpha/3^n = 2^{-n}(1-\alpha) + 2\cdot 3^{-n}(\alpha)$; so $x = \phi_\alpha(S)$, where $\phi_\alpha = (1-\alpha)\mu + \alpha\lambda$. Thus the map ϕ_α

takes $\mathbb{S}$ onto C_α, and it possesses nice properties. For instance, $\|\phi_\alpha - \phi_\beta\|_\infty = |\alpha - \beta|/6$; so the set $[x]$ of α's in $[0,1]$ for which $x \in C_\alpha$ is a closed subset of $[0,1]$. (If $d = d(x, C_\alpha) > 0$ then $x \notin C_\beta$ for $|\beta - \alpha| < 6d$.) Another relevant property of ϕ_α is that if $\alpha \neq \beta$ and $\varnothing \neq E \neq N$ then $\phi_\alpha(E) - \phi_\beta(E) = (\beta - \alpha)[\mu(E) - \lambda(E)] \neq 0$.

Define a linear ordering $<$ on $\mathbb{S}$ as follows: $E < F$ if there exists a positive integer n such that $E_{n-1} = F_{n-1}$ and $E_n \subsetneq F_n$, where $H_k = H \cap \{0,1,\dots,k\}$, $H \in \mathbb{S}$, $k \geqslant 0$ (i.e., $E < F \Leftrightarrow \phi_\alpha(E) < \phi_\alpha(F)$, $0 < \alpha \leqslant 1$).

Now we are ready to show that if $0 < x < 1$ then $[x]$ is a nowhere dense subset of $[0,1]$. Suppose $\alpha, \beta \in (0,1)$, $0 < x < 1$ and $\phi_\alpha(E) = x = \phi_\beta(F)$. Also, without loss of generality, suppose that $E < F$. Let n be the smallest positive integer in $F - E$. Let

$$G = E_n \cup \{n+1, n+2, \dots\}.$$

Then, for $0 < \gamma \leqslant 1$, $U_\gamma = (\phi_\gamma(G), \phi_\gamma(F_n))$ is a component of $[0,1] - C_\gamma$. Moreover, $\phi_\beta(G) < \phi_\beta(F_n) \leqslant x \leqslant \phi_\alpha(G) < \phi_\alpha(F_n)$. Thus, since U_γ deforms continuously from U_α to U_β as γ moves from α to β, there are γ's between α and β for which $x \notin C_\gamma$ (e.g., $x \notin C_\gamma$ when $0 < \phi_\gamma(F_n) - x < \inf\{\phi_\lambda(F_n) - \phi_\lambda(G);\ \lambda$ between α and $\beta\}$).

One of the referees of this note suggested using a nice subset K of the unit square to display the setting. To obtain K, draw line segments between points $(\lambda(E),1)$ and $(\mu(E),0)$, $E \in \mathbb{S}$, and let K denote the union of these intervals. One sees quickly that K is closed, that C_α is the intersection of K with the horizontal line $y = \alpha$, and that $[a]$ is the intersection of K with the vertical line $x = a$. Because the linear measure of C_α is $1 - \alpha$, some of those irrational Cantor sets are fat.

Reference

1. R. B. Darst, Some Cantor sets and Cantor functions, Math. Mag., 45 (1972) 2–7.

ON THE MONOTONICITY OF A CLASS OF EXPONENTIAL SEQUENCES

THOMAS P. DENCE

Mathematics Department, Bowling Green State University, Firelands Campus, Huron, OH 44839

It is well known that the sequence $(1 + 1/n)^n$ increases to e, whereas it is somewhat less familiar that the sequence $(1 + 1/n)^{n+1}$ decreases to e [**3**]. This note concerns the monotonicity of the sequence

$$a_n = (1 + 1/n)^{n+\alpha} \quad \text{for } 0 < \alpha < 1.$$

To this end, a sequence $\{\beta_k\}$ is defined by

$$\left[\frac{(k+1)^2}{k(k+2)}\right]^{\beta_k} = \left[\frac{k(k+2)}{(k+1)^2}\right]^{k+1}\left(\frac{k+1}{k}\right)$$

for $k = 1,2,\dots$. The value of β_k is precisely the value of α required for $a_k = a_{k+1}$. Several properties of $\{\beta_k\}$ will be essential.

LEMMA 1. *The sequence $\{\beta_k\}$ increases.*

Proof. Since

$$\beta_{k-1} = \frac{k\ln\left((k^2-1)/k^2\right) + \ln\left(k/(k-1)\right)}{\ln\left(k^2/(k^2-1)\right)},$$

we are led to consider the function $y = F(x)$ with

$$y = \frac{x\ln\left((x^2-1)/x^2\right) + \ln(x/(x-1))}{\ln\left(x^2/(x^2-1)\right)}, \qquad x > 1.$$

Its derivative is of the form $y' = \text{num}(x)/\text{den}(x)$ where $\text{den}(x) > 0$ for all $x > 1$, and $\text{num}(x)$ is given by

$$\frac{1}{x(x+1)}\ln\left(\frac{x^2}{x^2-1}\right) + \ln\left(\frac{x^2}{x^2-1}\right)\ln\left(\frac{x^2-1}{x^2}\right)$$
$$+\frac{2}{x^2-1}\ln\left(\frac{x^2-1}{x^2}\right) + \frac{2}{x(x^2-1)}\ln\left(\frac{x}{x-1}\right).$$

To show $\text{num}(x) > 0$ is equivalent to showing

$$\frac{1}{x(x+1)}\ln\left(\frac{x^2}{x^2-1}\right) + \frac{2}{x(x^2-1)}\ln\left(\frac{x}{x-1}\right) > \left[\ln\left(\frac{x^2}{x^2-1}\right)\right]^2 + \frac{2}{x^2-1}\ln\left(\frac{x^2}{x^2-1}\right).$$

This will be the case iff

$$\frac{-2}{x^2-1}\ln(x) + \frac{1}{x(x-1)}\ln(x+1) + \frac{1}{x(x+1)}\ln(x-1) > \left[\ln(x^2) - \ln(x^2-1)\right]^2.$$

Making the substitution $w = 1/x$, expanding $\ln(1+w)$ and $\ln(1-w)$ in series form, and multiplying through by $1 - w^2$ gives $w^4 + \frac{1}{6}w^6 + \frac{1}{15}w^8 + \cdots > (1-w^2)[\ln(1-w^2)]^2$. But, since the function $h(x) = x(1-x)^{-1/2} + \ln(1-x)$ is positive for $0 < x < 1$, this implies that $w^4 > (1-w^2)[\ln(1-w^2)]^2$; so the string of inequalities above is true. Consequently, $\{\beta_k\}$ is increasing since $\beta_k = F(k+1)$.

LEMMA 2. *The sequence* $\{\beta_k\}$ *converges to* $1/2$.

Proof. This result follows by applying L'Hospital's rule several times to $F(x)$.

Returning to the sequence $\{a_n\}$, we again examine the corresponding function of a continuous variable, namely, $f(x) = (1 + 1/x)^{x+\alpha}$, for $x \geqslant 1$. We are interested in knowing where the function is increasing and where it is decreasing. The derivative is

$$f'(x) = f(x)\left[\left(\frac{x+\alpha}{x+1}\right)\left(\frac{-1}{x}\right) + \ln\left(\frac{x+1}{x}\right)\right] = f(x)\,g(x).$$

Differentiating this function $g(x)$ gives

$$g'(x) = \frac{x(2\alpha-1)+\alpha}{\left[x(x+1)\right]^2}$$

from which we deduce that $g'(x) = 0$ when $x = \alpha/(1-2\alpha)$, which yields a positive x when $0 < \alpha < 1/2$.

Let us suppose that $1/2 \leqslant \alpha < 1$. Then $g'(x) > 0$, and since $g(1) < 0$ and $\lim_{x\to\infty} g(x) = 0$ then g is always negative. Consequently, f' is always negative; so f decreases. This establishes the following result.

LEMMA 3. *If* $1/2 \leqslant \alpha < 1$, *then* $\{a_n\}$ *is a decreasing sequence.*

This lemma can be found in Pólya and Szegö [**3**, vol. 1, p. 38, Problem 168], but we give an independent proof.

The value $1/2$ is the smallest value of x that will make $\{a_n\}$ decrease. Likewise, we shall show that there is a largest value (approximately 0.409) of a that will make $\{a_n\}$ increase.

If $0 < \alpha < 1/2$, then $x_c = \alpha/(1-2\alpha)$ is the only zero of g', and at that point $g(x_c) > 0$ and

also $g(x) > 0$ for all $x > x_c$. Since $\lim_{x\to\infty} g(x) = 0$ and $\lim_{x\to 0^+} g(x) = -\infty$, then we know that g has precisely one zero. Therefore f' has precisely one zero and f has precisely one critical point, call it x_0. Then, since $\lim_{x\to 0^+} f(x) = \infty$, $\lim_{x\to\infty} f(x) = e$, and $f'(x) > 0$ for all $x > x_c$, we know that f approaches the line $y = e$ from below; i.e., $f(x_0) < e$. So f is an increasing function for $x > x_0$ and a decreasing function for $x < x_0$. (See Fig. 1.) We can now determine the nature of $\{a_n\}$ for $0 < \alpha < 1/2$.

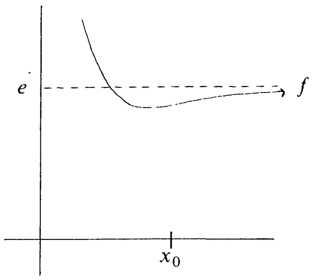


FIG. 1

LEMMA 4. *If* $\alpha = \beta_{k-1}$ *for some* k, *then* $a_1 > a_2 > \cdots > a_{k-1} = a_k < a_{k+1} < \cdots$.

Proof. Since $\alpha = \beta_{k-1}$ then $a_{k-1} = a_k$; so $f(k-1) = f(k)$, and thus $k - 1 < x_0 < k$.

The more interesting case is the following.

THEOREM. *If* $\beta_{k-1} < \alpha < \beta_k$, *then* $a_1 > a_2 > \cdots > a_{k-1} > a_k < a_{k+1} < a_{k+2} < \cdots$.

Proof. Since $a_k = (1 + 1/k)^{k+\alpha}$, then $a_{k-1} > a_k$ iff

$$\left(\frac{k}{k-1}\right)^{k-1+\alpha} > \left(\frac{k+1}{k}\right)^{k+\alpha}$$

iff

$$\left(\frac{k^2}{k^2-1}\right)^{\alpha} > \left(\frac{k^2-1}{k^2}\right)^{k}\left(\frac{k}{k-1}\right),$$

but

$$\left(\frac{k^2}{k^2-1}\right)^{\alpha} > \left(\frac{k^2}{k^2-1}\right)^{\beta_{k-1}} = \left(\frac{k^2-1}{k^2}\right)^{k}\left(\frac{k}{k-1}\right).$$

Thus $a_{k-1} > a_k$; so $k - 1 < x_0$, implying $a_1 > a_2 > \cdots > a_{k-1} > a_k$. Also $a_k < a_{k+1}$ iff

$$\left(\frac{k+1}{k}\right)^{k+\alpha} < \left(\frac{k+2}{k+1}\right)^{k+1+\alpha}$$

iff

$$\left[\frac{(k+1)^2}{k(k+2)}\right]^{\alpha} < \left[\frac{k(k+2)}{(k+1)^2}\right]^{k}\left(\frac{k+2}{k+1}\right),$$

but

$$\left[\frac{(k+1)^2}{k(k+2)}\right]^{\alpha} < \left[\frac{(k+1)^2}{k(k+2)}\right]^{\beta_k} = \left[\frac{k(k+2)}{(k+1)^2}\right]^{k+1}\left(\frac{k+1}{k}\right) = \left[\frac{k(k+2)}{(k+1)^2}\right]^{k}\left(\frac{k+2}{k+1}\right).$$

Thus $a_k < a_{k+1}$; so $x_0 < k + 1$, implying $a_k < a_{k+1} < a_{k+2} < \cdots$.

COROLLARY. *If* $0 < \alpha < \beta_1$, *then the sequence* $\{a_n\}$ *is increasing.*

Likewise, it follows that $\{a_n\}$ is decreasing when $1/2 \leqslant \alpha < 1$ since then $\beta_{k-1} < \alpha$ for all k. Generalizing yet further we see that $\{a_n\}$ is decreasing for $1 < \alpha$.

The reader may also wish to consult [**1**].

References

1. R. B. Darst, Comparison of estimates for e and e^{-1}, this MONTHLY, 86 (1979) 772–773.
2. N. S. Mendelsohn, An application of a famous inequality, this MONTHLY, 58 (1951) 563.
3. G. Pólya and G. Szegö, Problems and Theorems in Analysis, Springer, 1972, 1976.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

POLYNOMIALS IN TWO VARIABLES TAKING DISTINCT INTEGER VALUES AT LATTICE-POINTS

JOHN S. LEW
Mathematical Sciences Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

A polynomial $f(x,y)$ in real variables x,y, with real coefficients, defines a map $f\colon (x,y) \to f(x,y)$ from the Cartesian plane R^2 into the real line R. If sets N and Z respectively are the positive and signed integers, then Z^2 comprises the **lattice-points** of the Cartesian plane, that is, the points with integer coordinates. We review two problems, A and B: to find all real polynomials having respectively Properties A and B:

(A) the map f, restricted to N^2, takes this domain bijectively onto N;

(B) the map f, restricted to Z^2, takes this domain bijectively onto Z.

Problem A has a long history. Cauchy [**3**] used his now-standard method of diagonal enumeration to rewrite a double series $\Sigma_{i,j=1}^{\infty} a_{ij}$ as a single summation. Cantor [**1**] used this same scheme in a more fundamental way: to exhibit N^2 as a denumerable sequence. Indeed, Cantor observed that the polynomial

$$f^{\text{Cantor}}(x,y) = x + \tfrac{1}{2}(x+y-1)(x+y-2)$$

indexes the resulting sequence: if (x,y) is any point in N^2, then $f^{\text{Cantor}}(x,y)$ is its place in the enumeration. Thus Problem A, among its solutions, has at least **Cantor's pair** $f^{\text{Cantor}}(x,y)$ and its reflection $f^{\text{Cantor}}(y,x)$.

Fueter and Pólya [**6**], on Problem A, published an exchange of correspondence. Fueter, having read Carathéodory's exposition [**2**] of Cantor's results, wondered if any polynomial besides Cantor's pair indexed some enumeration of N^2. By studying the residue of an associated Dirichlet series, he excluded a class of quadratic polynomials. Pólya considered the area of the level set

$\{(x,y) \in R^2\colon 0 \leqslant x;\ 0 \leqslant y;\ 0 \leqslant f_d(x,y) \leqslant 1\}$, where f_d, for any polynomial f, is the homogeneous part of highest degree. By evaluating this area for the remaining quadratics, he excluded all such polynomials but Cantor's pair. Moreover, he ruled out polynomials of higher degree for which $f_d(x,y) > 0$ on the first quadrant except at (0,0).

Easily computable maps with Property A are **pairing functions** in recursive function theory [**12**]. Inductive use of such maps extends various concepts to finitely many integer variables. Hammer (written communication), in this context, posed Problems A and B in 1972. This MONTHLY published the latter [7], but it attracted no response.

Rosenberg [**13**], [**15**], independently formulated Problem A as a research question while investigating computer storage of multidimensional arrays. If a computer stores rectangular arrays by columns (respectively, rows) and an array gains further rows (respectively, columns) in the positive direction, then the computer must reshuffle storage locations for consistency. However, maps f from N^2 into N define **extendible schemes**, independent of such growth, for two-dimensional array storage [**14**], [**15**]. Indeed, points (x,y) in N^2 locate positions in a two-dimensional array, while integers $1,2,\ldots$ in N label cells in a computer memory—and any datum at position (x,y) takes the memory cell with **address** $f(x,y)$.

A "useful" map f should have nice additional properties. Data storage by **hashing** methods [**15**] allows many-to-one maps f, but each coincidence entails further programming steps. Thus f should not often repeat values. Also f should not often skip values, because this wastes memory space. Hence f should "approximate" a bijective map, and it should have a simple form. But bijections are the most obvious nearly bijective maps, and polynomials are the most obvious easily computable functions. A natural objective is therefore to find all bijective polynomials f—and this is precisely to solve Problem A. The Fueter-Pólya results prompt the conjecture that the Cantor pair are the only solutions. If these solutions fail some other demand, and if further evidence supports this conjecture, then extendible array storage must use either hashing methods or nonpolynomial maps.

Lew and Rosenberg [**10**] generalized the discussion of Problem A. If f is any map from R^2 into R, then f is a **packing** (respectively, **storing**) function on a subset S of R^2 whenever f, restricted to the lattice-points of S, maps them bijectively onto (respectively, injectively into) the nonnegative integers. Moreover f and S define a **density** $S \div f$; specifically,

$$S \div f = \lim_{n\to\infty} (1/n)\#\{Z^2 \cap S \cap f^{-1}([-n,+n])\}$$

when this limit is well defined. (Otherwise consider the lim sup and lim inf, respectively writing $S \mathbin{\overline{\div}} f$ and $S \mathbin{\underline{\div}} f$.) Here $S \cap f^{-1}([-n,+n])$ is a level set in the given S, and $\#\{\ldots\}$ is the cardinal number of contained lattice-points. Thus one calculates densities by a general theorem that counts lattice-points inside algebraic curves [**5**]. The definition extends a familiar number-theoretic density [**11**]. Moreover $S \div f$ equals Fueter's residue and Pólya's area whenever these yield nontrivial results.

A **rational sector** in the Cartesian plane is a closed sector, with nonvoid interior, having boundary rays with rational (or infinite) slope. The first quadrant, for example, is a rational sector; the whole plane, by convention, is another such. However, f has unit density on S whenever f is a packing function on S; and f is a packing function on the first quadrant precisely when $1 + f(x-1, y-1)$ is a function with Property A. Hence Lew and Rosenberg, for rational sectors S, seek all polynomial storing functions with unit density. Their results exclude all such functions of degrees 1, 3, and 4; and they exclude all **sectorially increasing** functions of higher degree. Here f is sectorially increasing on a domain S if S is a finite union of rational sectors S_i and f is eventually increasing on lattice points as $x^2 + y^2 \to \infty$ in each S_i. Also, they exclude all quadratic storing functions with unit density on rational half-planes, or the whole plane; and they exclude all such quadratics on the first quadrant except obvious translates of the Cantor pair.

The polynomial $f(x,y)$ has a singularity at infinity in a specified direction if its homogeneous part $f_d(x,y)$ is zero on the corresponding ray from the origin. It is these ***f*-singular** rays which

cause all the trouble. Originally, the author hoped that the values $f(x,y)$, for high-degree f, might be so sparse for large $|x|$, $|y|$ that $S \div f$ would be zero. But if

$$f(x,y) = ag(x) + b[y - c \cdot h(x)]^2, \deg(g) = 2, \deg(h) \geqslant 2,$$

where a,b,c are positive real numbers and g,h are monic polynomials, then [**10**]

$$(\text{first quadrant}) \div f = (\pi/2)(ab)^{-1/2} > 0;$$

while $\deg(f)$ can be any large even integer. Here the positive y-axis is an f-singular ray.

These results in two dimensions reinforce the conjecture that no polynomials but Cantor's pair have Property A. Also, the application to array storage motivates the analogous problem in m dimensions: to find all polynomials $f(x_1,\ldots,x_m)$, with real coefficients, which map N^m bijectively onto N. Thus, Chowla [**4**] has displayed one such polynomial for each positive m, and Lew [**9**] has constructed $c(m)$ essentially distinct such polynomials, where $(c(2), c(3), c(4), c(5),\ldots) = (1, 3, 11, 45,\ldots)$ is Sequence 1163 of [**16**], and [**8**],

$$c(m) \sim \text{constant} \cdot (3 + 2\sqrt{2})^m \cdot m^{-3/2} \quad \text{as } m \to \infty.$$

The constructed polynomials have least and greatest degrees m and 2^{m-1}.

These facts suggest a hybrid Problem C: to find a polynomial $f(x_1,\ldots,x_m)$ which maps Z^m surjectively onto N. This is clearly impossible when $m = 1$, and it is always possible when $m \geqslant 3$, by the famous theorem of Legendre and Gauss: each positive integer is the sum of three triangular numbers. Hence the two-dimensional case is the sole undecided one.

References

1. G. Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, J. Reine Angew. Math. 84 (1878) 242–258.

2. C. Carathéodory, Vorlesungen über reelle Funktionen, 1st ed., B. G. Teubner, Leipzig and Berlin, 1918, p. 29.

3. A. L. Cauchy, Cours d'analyse de l'École Royale Polytechnique, 1re partie: Analyse algébrique, l'Imprimerie Royale, Paris 1821, reprinted by Wissenschaftliche Buchgesellschaft, Darmstadt 1968, pp. 140–158.

4. P. Chowla, On some polynomials which represent every natural number exactly once, Norske Vid. Selsk. Forh., 34 (1961) 8–9; **MR** 30#4708.

5. H. Davenport, On a principle of Lipschitz, J. London Math. Soc., 26 (1951) 179–183; **MR** 13#323.

6. R. Fueter and G. Pólya, Rationale Abzählung der Gitterpunkte, Vierteljschr. Naturforsch. Gesellsch. Zurich, 58 (1923) 380–386.

7. F. D. Hammer, Advanced problem 6028*, this MONTHLY, 82 (1975) 410.

8. D. E. Knuth, The Art of Computer Programming, vol. 1: Fundamental Algorithms, Addison-Wesley, Reading, Mass., 1975, Exercise 2.21–12.

9. J. S. Lew, Polynomial enumeration of multidimensional lattices, Math. Systems Theory, 12 (1979) 253–270.

10. J. S. Lew and A. L. Rosenberg, Polynomial indexing of integer lattice-points I and II, J. Number Theory, 10 (1978) 192–214, 215–243; **MR** 57#18298a and b.

11. I. Niven and H. S. Zuckerman, An Introduction to the Theory of Numbers, 3rd ed., John Wiley, New York, 1972, p. 240.

12. H. Rogers, Jr., Theory of Recursive Functions and Effective Computability, McGraw Hill, New York, 1967, p. 64.

13. A. L. Rosenberg, Exploiting addressability in data graphs, in R. Rustin (ed.), Computational Complexity: Courant Computer Science Symposium I, Algorithmics Press, New York, 1973, 161–183.

14. ______, Allocating storage for extendible arrays, J. Assoc. Comput. Mach., 21 (1974) 652–670, **MR** 50#9037.

15. ______, Storage mappings for extendible arrays, in R. T. Yeh (ed.), Current Trends in Programming Methodology IV: Data Structuring, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp. 263–321.

16. N. J. A. Sloane, A Handbook of Integer Sequences, Academic Press, New York, 1973.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St Louis, MO 63121.

A NOTE ON AN ANALYTIC LOGARITHM

JOHN KLIPPERT
Department of Mathematics and Computer Science, James Madison University, Harrisonburg, VA 22807

One of the prettiest examples of the existence of an analytic logarithm is a problem (which I have not encountered in the texts) that I like to give to my classes in complex variables: characterize entire functions f and g such that $f^2 + g^2 = 1$. The "obvious" solution is that

(1) $f^2 + g^2 = 1$, f and g entire, if and only if

(2) $f = \cos(h)$ and $g = \sin(h)$ for some entire function h.

That (2) implies (1) is certainly clear. Going from (1) to (2) gives the student a nice challenge. For we must first write $(f + ig)(f - ig) = 1$. But then $f + ig \neq 0$. So we can find an entire function k such that $f + ig = e^k$. Then $f - ig = (f + ig)^{-1} = e^{-k}$, and these equations give $f = (e^k + e^{-k})/2$ and $g = (e^k - e^{-k})/2i$. Now

$$f = \cosh(k) = \cos(ik) = \cos(-ik) \text{ and } g = \frac{1}{i}\sinh(k) = \frac{1}{i}(-i\sin(ik)) = -\sin(ik) = \sin(-ik).$$

Finally, put $h = -ik$. Then h is entire, and $f = \cos(h)$ and $g = \sin(h)$.

The more general problem of finding meromorphic solutions to the equation $f^n + g^n = 1$, $n \geqslant 2$, is examined in [**1**].

Reference

1. Fred Gross, On the equation $f^n + g^n = 1$, Bull. Amer. Math. Soc., 72 (1966) 86–88.

AN ADVANCED CALCULUS PROOF OF THE FUNDAMENTAL THEOREM OF ALGEBRA

FRANK FORELLI
Department of Mathematics, University of Wisconsin, Madison, WI 53706

1. The purpose of this brief note is to set down a proof, which we think is attractive, of the fundamental theorem of algebra. This proof, which we have not seen elsewhere, uses just the elementary 'heory of (complex) differential forms and thus might be included (as an application of this theory) in what is nowadays a standard advanced calculus course.

2. Let

$$f(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_N z^N, \qquad \alpha_N \neq 0, \quad N > 0,$$

be a complex polynomial of positive degree N. It is to be proved that $f(z)$ has a complex root, i.e., that there is a complex number ζ such that $f(z) = 0$.

Suppose that $f(z) \neq 0$, $z \in \mathbb{C}$, and put

$$g(z,w) = w^N f(z/w) = \alpha_0 w^N + \alpha_1 z w^{N-1} + \cdots + \alpha_{N-1} z^{N-1} w + \alpha_N z^N.$$

Then $g(z,w) \neq 0$ if $(z,w) \neq (0,0)$. Denote by σ the real part of the complex variable w, $w = \sigma + i\tau$, and let Σ be the negative σ axis in $\mathbb{C}^2$, i.e.,

$$\Sigma = \{(0,\sigma) : \sigma \leq 0\}.$$

Now put $X = \mathbb{C}^2 \setminus \Sigma$. Then X is star-shaped with respect to the point $(0,1)$; hence on X the closed 1-form dg/g is exact, i.e., $dg/g = d\phi$ where $\phi \in C^1(X)$. (We might recall that ϕ may be defined by $\phi(z,w) = \int_\gamma dg/g$ where by γ we mean the linear path joining the point $(0,1)$ to (z,w), i.e., $\gamma(t) = (tz, 1 - t + tw)$, $0 \leq t \leq 1$.) We have $g(z,0) = \alpha_N z^N$; hence on the punctured plane $\{(z,0)\colon z \neq 0\}$ (which lies in X), $dg/g = N dz/z$. This gives

$$2\pi i = \int_{|z|=1} dz/z = (1/N)\int_{|z|=1,\, w=0} d\phi = 0.$$

Since this is false, we cannot have $f(z) \neq 0$ for all z, which proves that $f(z)$ has a root.

3. The idea of the foregoing is that everything is clear if we give ourselves more room in which to work. This is done by passing from $\mathbb{C}$ to $\mathbb{C}^2$.

The work for this paper was supported by the National Science Foundation.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Linear Transformation Fixed Scalar Multiple

S 22 [1979, 863]. *Proposed by Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada, and Roy Westwick, University of British Columbia.*

Let V and W be two vector spaces over the same field. Suppose f and g are two linear transformations $V \to W$ such that for every $x \in V$, $g(x)$ is a scalar multiple (depending on x) of $f(x)$. Prove that g is a scalar multiple of f.

Solution by Edward T. Wong, Oberlin College, Oberlin, Ohio. If $f = 0$ then $g = 0$. Suppose $f \neq 0$. Let $x \in V$ where $f(x) \neq 0$ and $g(x) = af(x)$. For any $y \in V$, if $f(y) = 0$ then $g(y) = af(y)$. If $f(y) \neq 0$ and $g(y) = bf(y)$, then for any scalar $d \neq 0$ there exists a c such that

$g(dx-y)=cf(dx-y)=cdf(x)-cf(y)=g(dx)-g(y)=adf(x)-bf(y)$. Thus, $(c-a)df(x)=(c-b)f(y)$. If $f(x)$ and $f(y)$ are linearly independent then $c=a=b$. If $f(x)$ and $f(y)$ are linearly dependent and $f(y)=df(x)$, then again $a=b$. Therefore $g=af$.

Also solved by 63 others, including the proposers.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by September 30, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2884. *Proposed by Lawrence Harris, University of Kentucky.*

Let $x_1,\dots,x_n$ be distinct real numbers. Set $S=\Sigma_1^n(1+x_k^2)^{n/2}/P(k)$, where $P(k)=\Pi_{j\neq k}|x_k-x_j|$. Prove $S\geqslant n$. When does $S=n$?

E 2885. *Proposed by T. Sekiguchi, University of Arkansas.*

Let T be a triangle. Construct the set of interior points of T at which the sum of the distances to the sides of T is equal to the arithmetic mean of the lengths of the altitudes of T.

E 2886. *Proposed by C. O. Oakley, Haverford College.*

Let $n>2$. How many tangent-normals can the graph of an nth-degree polynomial have? (A tangent-normal is a line that is tangent to the graph at one point and normal at another.)

E 2887. *Proposed by Arnold Adelberg, Grinnell College.*

Let $\mathcal{C}$ be a collection of subsets of the finite nonempty set U. (i) Characterize $\mathcal{C}$ if each proper subset of U meets an even number of sets in $\mathcal{C}$. (ii) Characterize $\mathcal{C}$ if each proper subset of U meets an odd number of sets in $\mathcal{C}$. (See E 2792 [1979, 702].)

E 2888. *Proposed by T. F. Mori and G. J. Szekely, Lorand Eötvös University, Hungary.*

Let $A_1, A_2,\dots,A_n$ be an arbitrary sequence of events in a probability space. Prove that

$$\prod_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{P(A_iA_j)}{P(A_i)P(A_j)}\right)\geqslant 1$$

always holds if the probabilities $P(A_i)$ are positive.

E 2889. *Proposed by I. J. Good, Virginia Polytechnic Institute.*

Let P be an arbitrary point in the plane of a regular polygon $A_1,A_2\dots A_n$. Let the foot of the perpendicular from P on line A_iA_{i+1} be Q_i (where A_{n+1} means A_1). Let x_i be $\pm$ length A_iQ_i: positive if Q_i, A_{i+1} are on the same side of A_i; negative otherwise. Prove that Σx_i is equal to half the perimeter of the polygon.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Block Matrix Not Equal to a Kronecker Product

E 2762 [1979, 223; 1980, 405]. *Proposed by Peter Hoffman, University of Waterloo, Canada.*

Let $A_1,\dots,A_n$ be $k\times k$ matrices over a field F, such that $A=A_1+\cdots+A_n$ is invertible. Show that the block-matrix

$$B = \begin{pmatrix} A_1 & A_2 & A_3 & \cdots & & A_n & 0 & \cdots & 0 \\ 0 & A_1 & A_2 & \cdots & & A_{n-1} & A_n & \cdots & 0 \\ \vdots & \vdots & \vdots & & & & & \vdots & \vdots \\ 0 & 0 & \cdots & A_1 & A_2 & & \cdots & \cdots & A_n \end{pmatrix}$$

has full rank, i.e., rank $(B) = mk$ where m is the number of block-rows.

Comment by Burkhard Schaffrin, University of Bonn, Germany. The solution [1980, 405] is incorrect, since B is not a Kronecker product as claimed.

Solution by M. L. J. Hautus, Eindhoven Institute of Technology, Eindhoven, Netherlands. Let $\mathbf{v} = (v_1,\ldots,v_m)$ be a (block) vector such that $\mathbf{v}B = 0 =$ the zero vector. (v_i is $1 \times k$.) Then

$$v_1A_1 = 0, v_1A_2 + v_2A_1 = 0,\ldots,v_mA_n = 0. \tag{*}$$

We pass now from the field F to the field $F(z)$ of rational functions over F. We set $\mathbf{v}(z) := \Sigma v_i z^i$, $A(z) := \Sigma A_i z^i$, summation being from 1 to n. Note that $A(z)$ is an invertible matrix of polynomials, since $\det A(1) = \det(A_1 + \cdots + A_n) \neq 0$. The equations (*) show that every power of z in the product $\mathbf{v}(z)A(z)$ has coefficient equal to the zero matrix. (The coefficient of z^2 is v_1A_1; the coefficient of z^3 is $v_1A_2 + v_2A_1,\ldots,$ the coefficient of z^{2n} is v_nA_n.) Hence $\mathbf{v}(z)$ is the zero vector, i.e., $v_1 = v_2 = \cdots = v_n = 0$.

F. S. Cater remarks that, if rank $(A_1 + \cdots + A_n) = r$, then at most $k - r$ linearly independent left vectors $v(z)$ exist satisfying (*), so that rank$B \geqslant rm$, and in fact rank$B = rm$.

Divisibility of $a^{2m} + b^{2m}$ by $a + b$

E 2772 [1979, 308]. *Proposed by R. B. McNeill, Northern Michigan University.*

Let m be a positive integer. Find all ordered pairs of positive integers (a,b) for which $(a + b)|(a^{2m} + b^{2m})$.

Solution by Lorraine L. Foster, CSU Northridge, Mark Merriman, Cambridge, Mass., and George Shulman, Teaneck, N.J. (independently). Let $(a,b) = (dA,dB)$ where $d = \gcd(a,b)$, $a = dA$, $b = dB$ and $\gcd(A,B) = 1$. Claim: $(a + b)|(a^{2m} + b^{2m}) \Leftrightarrow (A + B)|2d^{2m-1}$.

First note that, since $A^{2m} - B^{2m} \equiv 0 \pmod{A + B}$, we have

$$A^{2m} + B^{2m} \equiv 2A^{2m} \equiv 2B^{2m} \pmod{A + B}. \tag{1}$$

Hence,

$$\begin{aligned}(a + b)|(a^{2m} + b^{2m}) &\Leftrightarrow d(A + B)|d^{2m}(A^{2m} + B^{2m}) \\ &\Leftrightarrow (A + B)|d^{2m-1}(A^{2m} + B^{2m}) \\ &\Leftrightarrow (A + B)|d^{2m-1}\gcd(2A^{2m},2B^{2m}) \quad \text{from (1);} \\ &\Leftrightarrow (A + B)|2d^{2m-1}\end{aligned}$$

since $\gcd(A^{2m},B^{2m}) = 1$. This establishes the claim.

For a given pair of relatively prime positive integers A and B, the set S of integers d such that $(A + B)|2d^{2m-1}$ is determined as follows. Using the prime factorization of $(A + B)$, find the ***least*** positive integer D such that $(A + B)|2D^{2m-1}$. The set S then consists of all the multiples of D.

Finally, the desired solution set consists of all pairs $(a,b) = (kDA,kDB)$, where A,B are arbitrary relatively prime positive integers, k is any positive integer, and D is the least positive integer such that $(A + B)|2D^{2m-1}$.

Also solved by Stephen D. Bronn, Jeffrey Mitchell Cohen, Robert Gilmer, L. Kuipers (Switzerland), Man Kam Kwong, O. P. Lossers (Netherlands), Bernard L. Martin, M. R. Modak, M. P. Ojha, the University of South Alabama Problem Group, and the proposer.

Polynomials Divisible by $z^n \pm 1$

E 2810, E 2811 [1980, 60]; E 2819, E 2820 [1980, 137]. *Proposed by C. Notari, Mégrine-Coteaux, Tunisia.*

Let $T(z)$ be a polynomial with integral coefficients, having a common root with $P(z) = z^n - 1$. Supposing that for each root u_i of $P(z)$ we have $|T(u_i)| \leqslant 1$; prove that $T(z)$ is divisible by $z^n - 1$.

Let $T(z)$ be a non-constant polynomial with integral coefficients and $T(0) \neq 0$. If for each root u_i of $P(z)$ we have $|T(u_i)| \leqslant 1$, prove the existence of a unique integer k ($0 \leqslant k < n$) such that $T(z) + z^k$ or $T(z) - z^k$ is divisible by $z^n - 1$. Here $P(z) = z^n - 1$.

Let $T(z)$ be a polynomial with integral coefficients, having a root in common with $Q(z) = z^n + 1$. Supposing that for each root v_i of $Q(z)$ we have $|T(v_i)| \leqslant 1$, prove that $Q(z)$ divides $T(z)$.

Let $T(z)$ be a non-constant polynomial with integral coefficients and $T(0) \neq 0$. If for each root v_i of $Q(z) = z^n + 1$, we have $T(v_i) = 1$, prove the existence of a unique integer k ($0 \leqslant k < n$), such that $T(z) + z^k$ or $T(z) - z^k$ is divisible by $z^n + 1$.

Solution by A. A. Jagers, Technische Hogeschool Twente, Enschede, Netherlands. All four problems immediately follow from the following proposition (note that E2811 and E2820 contain some inaccuracies).

PROPOSITION. Let $T(z)$ be a polynomial with integral coefficients such that $|T(u_i)| \leqslant 1$ for all roots u_i of $P(z)$ where $P(z) = z^n - 1$ or $P(z) = z^n + 1$. Then either $T(z)$ is divisible by $P(z)$ or there exists a unique integer k, $0 \leqslant k < n$, such that $T(z) + z^k$ or $T(z) - z^k$ is divisible by $P(z)$.

Proof. By polynomial long division we have $T(z) = A(z)P(z) + B(z)$ where both $A(z)$ and $B(z)$ are polynomials with integer coefficients and $B(z)$ has degree $\leqslant n - 1$. Suppose $B(z)$ has a zero of order k at $z = 0$, $B(z) = z^k C(z)$, say, with $C(0) \neq 0$ (if $B(z) \equiv 0$, we are done). Then $|C(u_i)| = |B(u_i)| = |T(u_i)| \leqslant 1$, but, e.g., by Lagrange interpolation with respect to the u_i,

$$C(0) = \pm n^{-1} \sum_{i=1}^{n} C(u_i),$$

where the sign depends on the choice of $P(z)$. Hence, by a triangle inequality,

$$0 < |C(0)| = |n^{-1} \sum_{i=1}^{n} C(u_i)| \leqslant n^{-1} \sum_{i=1}^{n} |C(u_i)| \leqslant 1.$$

However, $C(0)$ is an integer; so either $C(0) = 1$ or $C(0) = -1$, but then necessarily $C(u_i) = 1$ for all i or $C(u_i) = -1$ for all i. That is, $C(z) \equiv 1$ or $C(z) \equiv -1$, because the degree of $C(z)$ is less than n. □

Also solved by W. A. Al-Salam and A. Meir (Canada), Miroslav D. Ašić (Yugoslavia), B. L. Aubertin (Canada) & Peter Borwein (England), Kenneth L. Bernstein, Amer Bešlagić (Yugoslavia), Theodore Bolis, William Bosch, J. E. Chance, Chico Problem Group, Ronald Evans, T. Fujinawa, Enzo R. Gentile, Michael Golomb, L. Kuipers (Switzerland), Kunte Kinte AAMS Study Group, O. P. Lossers (Netherlands), L. E. Mattics, M. R. Modak & S. A. Katre (India), Dalton Orr, Jens Schwaiger (Austria), David A. Singer, St. Olaf College Problem Group, A. Tyszka (Poland), Paul Zwier, and the proposer.

Fermat's Theorem for an Infinity of Exponents

E 2812 [1980, 60]. *Proposed by B. J. Powell, Kirkland, Washington.*

Prove that for every odd prime p, there exists an infinite set of pairwise relatively prime

integers n such that the equation $x^{np}+y^{np}=z^{np}$ has no solution in positive integers x,y,z with $xyz \not\equiv 0 \bmod p$).

[This is an extension of the proposer's theorem in "Proof of a Special Case of Fermat's Last Theorem," vol. 85, pp. 750–751, this MONTHLY, November 1978.]

Solution by L. E. Mattics, University of South Alabama. We replace the hypothesis $x,y,z>0$ by $xyz \neq 0$. For $p=3$ Fermat's conjecture holds so $x^{3n}+y^{3n} \neq z^{3n}$ if $xyz \neq 0$. If $p>3$, by Dirichlet's theorem, there is a prime $q=3+4pt$ with $q>p+1$. Let $n=n_1=(q-1)/2$ and suppose that $x^{np}+y^{np}=z^{np}$ with $(xyz,p)=(x,y,z)=1$ and $xyz \neq 0$. Modulo q, each of x^n,y^n,z^n is 0, -1, or 1; so one of x^n,y^n,z^n is 0 modulo q. Since np is odd we may assume that $z \equiv 0 \pmod q$. It is well-known that if $(r,s)=(r+s,p)=1$ then $(r+s,(r^p+s^p)/(r+s))=1$; so $(x^n+y^n,(x^{np}+y^{np})/(x^n+y^n))=1$ and $x^n+y^n=a^{np}$, $(x^{np}+y^{np})/(x^n+y^n)=b^{np}$ with $(a,b)=1$. Since p is odd, $(x,y)=1$ and x^n,y^n are ± 1, $x^{np}+y^{np} \equiv x^n+y^n \pmod q$ so $a \equiv 0 \pmod q$.

By the binomial theorem, $x^{np}+y^{np}=x^{np}+(a^{np}-x^n)^p=-(x^n)^{p-1}pa^{np}+(a^{np})^2A=(ab)^{np}$ for some integer A; therefore, $b^{np}=(x^n)^{p-1}p+a^{np}A$ and so $p \equiv \pm 1 \pmod q$. But this implies that $q \leqslant p+1$, which contradicts the choice of q. Hence there is at least one such odd number n_1. Suppose that we have found m such coprime numbers $n_1,n_2,\ldots,n_m$ such that $x^{n_ip}+y^{n_ip}=z^{n_ip}$ has no integral solutions x,y,z with $(xyz,p)=(x,y,z)=1$ and $xyz \neq 0$. If $(n_1n_2\cdots n_m,3)=1$, there is a prime $q=3+4pn_1\cdots n_mt$ with $q>p+1$. If (say) $3^a \| n_1$ with $a>0$ we use the Chinese remainder theorem to find an r such that $r \equiv 3 \pmod{4\cdot 3^{-a}p\cdot n_1\cdots n_m}$ and $r \equiv 2 \pmod 3$; then Dirichlet's theorem assures us of a prime $q \equiv r \pmod{4\cdot 3^{-a+1}p\cdot n_1\cdots n_m}$ with $q>p+1$. In either case set $n_{m+1}=(q-1)/2$; then $(n_1n_2\cdots n_m,n_{m+1})=1$, and we repeat the argument above as for n_1.

The proposer draws the following corollary.

THEOREM. *There are infinitely many pairwise relatively prime positive integers n, such that there is a prime p dividing n for which $x^n+y^n=z^n$ has no solutions in positive integers x,y,z, $p \nmid xyz$.*

Proof. Pick $q_1>p_1+1$, as in the solution above, and set $n_1=p_1(q_1-1)/2$, and conclude as in the solution above that $x^{n_1}+y^{n_1}=z^{n_1}$ has no solutions with $p_1 \nmid xyz$. Then pick $q_2 \equiv -1 \bmod n_1$, and $n_2=p_2(q_2-1)/2$, with $q_2>p_2+1>p_2>p_1$, etc. This generates an infinite set of integers $n_1,n_2,\ldots$ all pairwise relatively prime.

Also solved by Robert Breusch and Robert E. Shafer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by September 30, 1981. The solver's full post-office address should be on each sheet.

6344. *Proposed by M. Machover, St. John's University.*

Let $\theta_1,\ldots,\theta_n$ be distinct angles (mod 2π). Let $A_1,\ldots,A_n$ be nonzero complex numbers. Let g be a function defined on the open interval (a,b) such that $(x,g(x))$ describes a continuous simple arc as x ranges over (a,b). Consider the sum

$$S(x)=\sum_1^n A_j\exp\big(i\big((\sin\theta_j)x+(\cos\theta_j)g(x)\big)\big).$$

(a) For each even integer $n>2$ find A's, θ's, g, and (a,b) such that $S(x)=0$ for all $x\in(a,b)$.

(b) Show that, if $n = 3$, then there is no choice of A's, θ's, g and (a,b) such that $S(x) = 0$ for all $x \in (a,b)$.

(c)* Conjecture: if n is odd, then there is no choice of A's, θ's, g and (a,b) such that $S(x) = 0$ for all $x \in (a,b)$.

6345. *Proposed by Gary L. Walls, University of Southern Mississippi.*

Which groups have exactly 6 subgroups?

6346. *Proposed by M. Slater, University of Bristol, England.*

Define the sets A_n, B_n of integers ($n = 1,2,\ldots$) recursively as follows: $A_1 = \phi$, $B_1 = \{0\}$, $A_2 = \{1\}$, $B_2 = \{0\},\ldots$. In general, each element of A_{n+1} is obtained by adding one unit to each element of B_n. The set $B_{n+1} := A_n \cup B_n - A_n \cap B_n$. Prove that $B_n = \{0\}$ if and only if n is a power of 2.

SOLUTIONS OF ADVANCED PROBLEMS

Complete Categorical Theories

6272 [1979, 509]. *Proposed by P. Olin, York University, and Kenneth W. Smith, University of Toronto.*

It is known (Waszkiewicz and Weglorz, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys., 17 (1969) 195–199; see top of page 197) that there is a complete, $\aleph_1$-categorical theory T of first-order logic such that the direct product $T \times T$ is not $\aleph_1$-categorical. Are there complete first-order theories T_1, T_2 with T_1 the theory of a finite model, T_2 $\aleph_1$-categorical, and $T_1 \times T_2$ not $\aleph_1$-categorical? If so, find such a pair T_1, T_2 with the cardinality of the model of T_1 as small as possible.

Solution by the proposers. Yes, such theories exist, with T_1 the theory of a one-element model and T_2 categorical in all infinite cardinalities. Let L be the first-order language with identity as a logical symbol and with P a one-place predicate symbol and R a two-place predicate symbol. Let T_1 be the theory of the model $\mathfrak{U}$ having universe $\{a\}, P(a)$ true and $R(a,a)$ false. Let T_2 be the theory axiomatized by

$$(\forall x)(\forall y)(R(x,y) \to P(x) \wedge \sim P(y)),$$
$$(\forall x)(P(x) \to (\exists! y) R(x,y))$$
$$(\forall y)(\sim P(y) \to (\exists! x) R(x,y))$$

and the sentences which imply that all models of T_2 are infinite. Note that if $\mathfrak{B}$ is any model of T_2, $R^{\mathfrak{B}}$ provides a bijection from $P^{\mathfrak{B}}$ to $(\sim P)^{\mathfrak{B}}$. It follows that T_2 is categorical in all infinite cardinalities. Let $\mathfrak{B}$ be a model of T_2 of cardinality $\aleph_1$.

Let $\mathfrak{B}_1$ be a model with universe of cardinality $\aleph_1, P^{\mathfrak{B}_1}$ of cardinality $\aleph_0$ and $R^{\mathfrak{B}_1}$ empty. It is not hard to see (for example, by the method of games as in Monk, *Mathematical Logic*, Springer-Verlag, 1976, Theorem 26.14) that $\mathfrak{B}_1$ and $\mathfrak{U} \times \mathfrak{B}$ have the same complete theories. And $\mathfrak{B}_1 \not\cong \mathfrak{U} \times \mathfrak{B}$ because $P^{\mathfrak{B}_1}, P^{\mathfrak{U}\times\mathfrak{B}}$ have cardinalities $\aleph_0, \aleph_1$, respectively.

This solution can be compared with the method of proof of Proposition 3.2.11(ii) (and can be contrasted with Exercise 7.1.2) in Chang and Keisler, *Model Theory*, North-Holland, 1973.

If either the language L contains only one nonlogical symbol or if L is countable monadic, and

if T_1 is the theory of a one-element model, then it can be shown that if T_2 is $\aleph_1$-categorical so is $T_1 \times T_2$. Hence the example above is, in a sense, best possible.

On the other hand, A. Mekler has pointed out the following example. Suppose the language L contains only one nonlogical symbol P, a one-place predicate symbol. Let n be finite, $n > 1$, and let T_1 be the theory of the model of cardinality n all but one of whose elements satisfies P. Let T_2 be the theory axiomatized by $(\forall x)P(x)$ together with the axioms which imply that all models of T_2 are infinite. T_2 is categorical in all infinite cardinalities. But $\mathfrak{B}$ is a model of $T_1 \times T_2$ iff both $P^{\mathfrak{A}}$ and $(\sim P)^{\mathfrak{A}}$ are infinite. So $T_1 \times T_2$ is not $\aleph_1$-categorical.

A False Criterion for Continuity

6273 [1979, 596]. *Proposed by K. L. Chung, Stanford University.*

Let f be a real-valued function defined on $(-\infty, +\infty)$ and continuous from the right everywhere. Suppose also that the following is true:

$$\lim_{n\to\infty}\left[\max_{-\infty<k<\infty}\left|f\left(\frac{k+1}{n}\right)-f\left(\frac{k}{n}\right)\right|\right]=0$$

where n and k are integers, $n \geqslant 1$. Is f continuous in $(-\infty, +\infty)$?

Solution by F. S. Cater, Portland State University. We prove that such a function need not be everywhere continuous. Indeed, we construct such a function which is unbounded in every neighborhood of 0. (We switch left and right for convenience, but that does not matter.)

Let $n_0 = 1$, $n_1 = 5$, and in general for each $j > 0$ let n_j be the smallest integer $> n_{j-1}$ such that $n_{j-1}^{-1} + (n_{j-1}+1)^{-1} + \cdots + (n_j - 1)^{-1} \geqslant 2j$.

By a *spike* with support (a,b), we mean the function v defined by

$$v(x)=\begin{cases}0 & \text{for } x \notin (a,b)\\ 1 & \text{for } x=\tfrac{1}{2}(a+b)\\ 2(x-a)(b-a)^{-1} & \text{for } a<x<\tfrac{1}{2}(a+b)\\ 2(b-x)(a-b)^{-1} & \text{for } \tfrac{1}{2}(a+b)<x<b.\end{cases}$$

For any positive integer n and any function h, let

$$W(n,h)=\sup_{-\infty<k<\infty}|h((k+1)/n)-h(k/n)|.$$

For each integer $j>0$, let $g(n_j)$ be the spike whose support is the interval $(2^{-j},2^{1-j})$. Let $g(n_j+1)$ be the spike whose support is the longest interval with left endpoint = midpoint of the support of $g(n_j)$, with length $\leqslant \frac{1}{4}$ length of the support of $g(n_j)$, and containing no point of the form k/n for any n satisfying $W(n,g(n_j)) \geqslant 2^{-j}$. Let $g(n_j+2)$ be the spike whose support is the longest interval with left endpoint = midpoint of the support of $g(n_j+1)$, with length $\leqslant \frac{1}{4}$ length of the support of $g(n_j+1)$, and containing no point of the form k/n for any n satisfying $W(n,g(n_j)) \geqslant 2^{-j}$ or $W(n,g(n_j+1)) \geqslant 2^{-j}$. Finally, let $g(n_{j+1}-1)$ be the spike whose support is the longest interval with left endpoint = midpoint of the support of $g(n_{j+1}-2)$, with length $\leqslant \frac{1}{4}$ length of the support of $g(n_{j+1}-2)$, and containing no point of the form k/n for any n satisfying

$$W(n,g(n_j)) \geqslant 2^{-j} \text{ or } W(n,g(n_j+1)) \geqslant 2^{-j} \text{ or } \cdots \text{ or } W(n,g(n_{j+1}-2)) \geqslant 2^{-j}.$$

Now $g(k)$ is defined for all $k \geqslant 1$.

Put $f_j = n_j^{-1}g(n_j) + (n_j+1)^{-1}g(n_j+1) + \cdots + (n_{j+1}-1)^{-1}g(n_{j+1}-1)$ and $f = \sum_{j=1}^{\infty} f_j$. Then it follows that

$$W(n,f_j) \leqslant n_j^{-1}W(n,g(n_j)) + (n_j+1)^{-1}W(n,g(n_j+1)) + \cdots + (n_{j+1}-1)^{-1}g(n_{j+1}-1))$$
$$\leqslant n_j^{-1} + (2j+3)2^{-j}.$$

Now the functions $\Sigma_{j=1}^{J-1}f_j, f_J, f_{J+1}, f_{J+2}, \ldots$ have mutually disjoint supports; so

$$W(n,f) \leqslant \sup\left[W\left(n,\Sigma_{j=1}^{J-1}f_j\right), W(n,f_J), W(n,f_{J+1}), W(n,f_{J+2}),\ldots\right]$$
$$\leqslant \sup_{j\geqslant J}\left[W\left(n,\Sigma_{j=1}^{J-1}f_j\right), n_j^{-1} + (2j+3)2^{-j}\right)\Big]$$
$$\leqslant \sup\left[W\left(n,\Sigma_{j=1}^{J-1}f_j\right), n_J^{-1} + (2J+3)2^{-J}\right)\Big].$$

But $\Sigma_{j=1}^{J-1}f_j$ is continuous; so for large enough n, $W(n,f) \leqslant (2J+3)2^{-J} + n_J^{-1}$. Of course the choice of J is arbitrary; so $\lim_{n\to\infty}W(n,f) = 0$.

The function f is left continuous because each $x \in R$ is the right endpoint of an interval that meets the support of at most finitely many f_j. On the other hand,

$$\sup f_j \geqslant \tfrac{1}{2}n_j^{-1} + \tfrac{1}{2}(n_j+1)^{-1} + \cdots + \tfrac{1}{2}(n_{j+1}-1)^{-1} \geqslant j$$

and every neighborhood of 0 contains the supports of infinitely many f_j. Thus f is unbounded in every neighborhood of 0, and f is discontinuous at 0.

REMARK. $F = \min(1,f)$ is a bounded function of the kind considered, discontinuous at 0. Then $\Sigma_n 2^{-n}F(x + r_n)$, where the sequence (r_n) is an enumeration of the rational numbers, is such a function discontinuous at each rational point. Moreover, no function can be left (right) continuous and right (left) discontinuous at uncountably many points.

Also solved by L. E. Mattics and J. Michael Steele. Steele's solution is contained in his paper *A counterexample related to a criterion for a function to be continuous*, Proc. Amer. Math. Soc., 79 (1980) 107–109.

Isometry Groups

6275 [1979, 597]. *Proposed by S. Foldes and E. Howorka, Grenoble, France, and the University of Florida.*

Let r be a metric on R^n giving the same topology as the usual Euclidean metric d. Let $I(r), I(d)$ denote their groups of isometries. S. Ulam conjectured recently that, if $I(r)$ contains an isomorphic copy of $I(d)$, then $I(r) \cong I(d)$. This conjecture has not been settled yet. Show that $I(d) \subseteq I(r)$ implies $I(r) = I(d)$.

Solution by the proposers. Observe first that the Euclidean isometry group is distance-transitive, and, therefore, $r(x,y)$ depends only on $d(x,y)$ and not on the particular choice of x and y. Moreover,

($*$) For every $\delta > 0$, there exist $\delta_1, \delta_2 > 0$ such that $\delta_1 < \delta$, and, for each pair of points x,y,

$$d(x,y) = \delta_1 \Rightarrow r(x,y) = \delta_2$$
$$d(x,y) > \delta_1 \Rightarrow r(x,y) > \delta_2.$$

Proof of ($*$). By the initial remark, we may assume without loss of generality that x is any point of R^n. Since d and r are topologically equivalent, there is a δ', $0 < \delta' < \delta$, such that $r(x,z) \leqslant \delta'$ implies $d(x,z) < \delta$ for each z. The set $B = \{z\colon r(x,z) \leqslant \delta'\}$ is compact; let z_0 have the maximal d-distance from x among all elements of B, and set $\delta_1 = d(x,z_0)$, $\delta_2 = r(x,z_0)$.

Let f be an arbitrary r-isometry of R^n, and let x,y be elements of R^n. We shall show that for every $\epsilon > 0$,

$$d(f(x),f(y)) < d(x,y) + \epsilon.$$

Applying this to f^{-1} and letting $\epsilon \to 0$, we obtain $d(x,y) = d(f(x),f(y))$, and the main assertion follows immediately.

Consider an arbitrary $\epsilon > 0$, and let f,x,y be as above. Clearly, the map $z \mapsto d(f(z),f(y))$ is continuous. Hence, there is a δ_0 such that $d(z,y) < \delta_0$ implies $d(f(z),f(y)) < \epsilon$. Let δ_1,δ_2 be as in $(*)$ applied to $\delta = \delta_0$. Let k be the largest nonnegative integer with $k\delta_1 \leqslant d(x,y)$. Let $x_0,x_1,\ldots,x_k$ be points on the straight line segment joining x and y such that $d(x,x_i) = i\delta_1$, $0 \leqslant i \leqslant k$. Thus, $d(x_i,x_{i+1}) = \delta_1$ for each $i \leqslant k-1$. Since f is an r-isometry, by $(*)$ we must have $d(f(x_i),f(x_{i+1})) \leqslant \delta_1$ for each $i \leqslant k-1$. On the other hand, $d(x_k,y) \leqslant \delta_0$ implies $d(f(x_k),f(y)) < \epsilon$. It follows that

$$\begin{aligned} d(f(x),f(y)) &\leqslant \textstyle\sum_{i=0}^{k-1} d(f(x_i),f(x_{i+1})) + d(f(x_k),f(y)) \\ &< k\delta_1 + \epsilon \\ &\leqslant d(x,y) + \epsilon. \end{aligned}$$

This completes the proof.

Also solved by Andrew Vogt. F. S. Cater generalized the result to any inner product space. The proposers' solution also establishes this generalization with virtually no change.

Groups Generated by Screw Motions

6276 [1979, 709]. *Proposed by R. K. Oliver, Pittsburgh, Pa.*

Let g and h be two screw motions of Euclidean three-space with positive angles less than $\pi/3$ and nonparallel axes. Show that the group generated by g and h is not discrete.

Solution by the proposer. Let the origin be a point on the axis of g, let (g_k) be the sequence of transforms $g_1 = hgh^{-1}$, $g_2 = g_1gg_1^{-1}$, $g_3 = g_2gg_2^{-1}$, etc., and let A,A_k and a,a_k be the orthogonal and translative parts of g,g_k. Thus $gx = Ax + a$, $g_kx = A_kx + a_k$ $(x \in R^3)$. Further, let α denote the angle of A (= angle of g) and β_k the angle between the axes of A and A_k. Then, since A_k rotates the axis of A into the axis of A_{k+1} and the angle of A_k is α, we have $\sin\frac{1}{2}\beta_{k+1} = \sin\frac{1}{2}\alpha \,\sin\beta_k$. Hence $\sin\beta_{k+1} < m\sin\beta_k$, where $m = 2\sin\frac{1}{2}\alpha < 1$. Thus (roughly speaking) the axes of the A_k spiral in to the axis of A. Hence the A_k and therefore also the g_k are distinct. Now let K be a ball about the origin with the following two properties: (1) Any transform of g whose axis meets K maps the origin to a point of K. (2) The axis of g_1 meets K. (Since $m < 1$, such a ball clearly exists.) Then, since the axis of g passes through the origin, it follows by induction that the axis of every g_k meets K. Hence the a_k and therefore also the g_k are bounded. Thus, since the g_k are distinct, the group generated by g and h contains a bounded infinite set, and so is not discrete.

REMARKS. Proofs of results essentially equivalent to the one of this problem have been given by Rohn [3], [4], Schönflies [5], [6], and recently by Delone and Štogrin [1]. The preceding solution is a simplification of the proofs of Rohn and has certain advantages over the proof of Delone and Štogrin. This problem is interesting because of its relation to crystallography and because its n-dimensional analogue is one of the main parts of Hilbert's 18th problem on crystallographic groups (cf. [2] and the proposer's paper *On Bieberbach's analysis of discrete Euclidean groups*, submitted for publication).

References

1. B. N. Delone and M. I. Štogrin, Simplified proof of the Schönflies theorem (Russian), Dokl. Akad. Nauk SSSR, 219 (1974) 95–98. English translation: Sov. Phys. Dokl., 19 (1974) 727–729. MR 51 (1976) #6556.

2. J. Milnor, Hilbert's Problem 18: On crystallographic groups, fundamental domains, and on sphere packing, Proc. Symp. Pure Math. 28, Amer. Math. Soc., Providence, R.I., 1976, pp. 491–506. MR 55 (1978) #3108.

3. K. Rohn, Krystallstruktur und regelmässige Punktgruppen, Ber. Sächs. Akad. Wiss., 51 (1899) 445–455.

4. K. Rohn, Einige Sätze uber regelmässige Punktgruppen, Math. Ann., 53 (1900) 440–449.

5. A. Schönflies, Krystallsysteme und Krystallstructur, Leipzig, 1891.

6. ______, Theorie der Kristallstruktur, Berlin, 1923.

An Explicit Map $(0,1) \cap Q \simeq [0,1] \cap Q$

6282* [1979, 869]. *Proposed by David P. Robbins, Hamilton College.*

It is known (Wilansky, *Topology for Analysis*, p. 113) that any two countable metric spaces with no isolated points are homeomorphic. Give an explicit homeomorphism between the space of rational numbers r with $0 < r < 1$ and that of rationals t with $0 \leqslant t \leqslant 1$.

Solution by A. Meir and J. Muldowney, Edmonton, Alberta. Let α be any irrational number, $0 < \alpha < 1$, and let $0 = r_0 < r_1 < r_2 < \cdots$ be rationals such that $r_n \to (1-\alpha)/4$. We define the mapping f from the rationals in $[0,1]$ onto the rationals in $(0,1)$ by

$$f(0) = \frac{1}{4}$$

$$f(t) = \frac{1}{4} - 2^{k-1}t, \qquad \text{if } \alpha \cdot 2^{-2k-2} < t < \alpha \cdot 2^{-2k-1}, \qquad k = 0,1,2,\ldots$$

$$f(t) = \frac{1}{4} + 2^{k-1}t, \qquad \text{if } \alpha \cdot 2^{-2k-1} < t < \alpha \cdot 2^{-2k}, \qquad k = 1,2,\ldots$$

$$f(t) = \frac{1}{2} - t - r_{k-1} - r_k, \quad \text{if } \frac{1+\alpha}{4} - r_k < t < \frac{1+\alpha}{4} - r_{k-1}, \quad k = 1,2,\ldots$$

$$f(t) = t, \qquad \text{if } \frac{1+\alpha}{4} < t \leqslant \frac{1}{2},$$

$$f(t) = 1 - f(1-t), \qquad \text{if } \frac{1}{2} < t \leqslant 1.$$

It is not difficult to see that f is a homeomorphism.

Also solved by Kenneth P. Bube, F. S. Cater, Chico Problem Group, G. A. Heuer & Karl Heuer, Robert B. Israel, Joel Levy, Bernard J. McCabe, Curt McMullen, Barbara L. Osofsky, Nicholas Passell, Pavel Pyrih (Czechoslovakia), Charles Riley, David M. Wells, and David J. Wright.

REVIEWS

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

Books submitted for review should be sent to Professor John H. Ewing, Department of Mathematics, Indiana University, Bloomington, IN 47405.

Finite Mathematics: A Modeling Approach. By Marvin L. Bittinger and J. Conrad Crown. Addison-Wesley, Reading, Massachusetts, 1977. xi + 375 pp. + Appendices and Index. \$15.95. (Telegraphic Review, October 1977.)

Finite Mathematics. By Daniel P. Maki and Maynard Thompson. McGraw-Hill, New York, 1978. x + 452 pp. $14.95. (Telegraphic Review, August-September 1978.)

Since its inception at Dartmouth in the mid-1950's, the finite mathematics course has become a popular vehicle for exposing students to applications of mathematics in diverse fields. This course has induced thousands of college students who might otherwise be "mathematically timid" to take another look at our discipline. Many of these students evidently like what they see in the course; they often go on to study calculus, statistics, or another area of mathematics. For other students who do not major in the sciences or the quantitative social sciences, the finite mathematics course may terminate a formal study of mathematics. We believe that both groups of students can be well served by their encounter with finite mathematics and its applications.

Over the past four years, we have taught courses in finite mathematics using the texts under review as well as the newest edition of the modern classic by Kemeny, Snell, and Thompson (extended review, June-July 1975). Unlike the earlier editions of Kemeny, Snell, and Thompson, these three texts are well suited to a course that moves quickly into discrete probability, Markov chains, and the beginnings of statistics. All three include discussions of set theory and counting, matrix algebra and linear equations, linear programming and the simplex method, and game theory and networks. The traditional chapter on logic and truth tables is relegated to an appendix in Bittinger and Crown and deleted altogether from Maki and Thompson. Neither of the more recent texts contains a chapter on computer programming. Maki and Thompson concludes with a chapter (which we did not use) on evaluating investment options.

Because the text by Maki and Thompson is more akin to the familiar text of Kemeny, Snell, and Thompson, we consider it first. Although we judged it the least physically attractive of these texts, we found it appealing in other ways. It is written with the high standards of mathematics and pedagogy that typify the more advanced and pioneering text on mathematical models by the same authors.

One strength of this text is the flexibility it provides the instructor. After an introductory chapter on mathematical models, sets, and functions, it is organized in three main units: probability models, linear models, and applications. The first two units are independent and self-contained. Four chapters on applications are also independent of each other and require as background only the core chapters of the first two units. A user of this text can thus find material for two nonoverlapping courses, neither of which would require the other as prerequisite.

The authors' care in developing a sound understanding of important concepts is illustrated by their exposition of Bayes's Theorem. They first derive the result through an illustration involving audio-tape purchases by the Mul T. Decibel Music Company. A tree diagram clarifies the structure of the argument. Only then do they present the more rigorous development, based on repeated application of the definition of conditional probabilities. The close parallels between these developments are emphasized. A subsequent illustration with an influenza epidemic extends the development to three stages; our students appreciated Maki and Thompson's frequent discussions of nontrivial and realistic applications.

The normal approximation to a binomial distribution is well motivated and uses illustrations effectively. We prefer Maki and Thompson's development to that of Kemeny, Snell, and Thompson; we think that its use of the continuity correction gives students more insight into the Central Limit Theorem. The degree of accuracy of the approximation is explored and illustrated. (One complaint: The continuity correction was not employed when answers to the exercises were obtained.) In general, we believe that our students gained a fuller understanding of normal distributions from this text than from the others.

Unlike many of the 40 finite mathematics texts listed in 1978 and 1979 in the Telegraphic Reviews of this MONTHLY, the Maki and Thompson text will survive to a second edition. We offer four suggestions for the revision:

1. The discussion of decision problems in the final section of the chapter on statistics could benefit from revision. The discussion of composite hypotheses and sample size determination is rather formidable for students who do not understand type I and type II error

probabilities in the simplest settings. We prefer the discussion of hypothesis testing found in Kemeny, Snell, and Thompson.

2. We would prefer that a matrix solution of systems of linear equations be given using only the elementary row operations. The use of a partial reduction, followed by a "back substitution" in the resulting equations fails to encourage our freshmen to leave substitution behind and focus on the powerful elimination technique. We have encountered no pedagogical difficulties in using the full reduction; it also enables us to introduce our students to an interactive computer program for fully reducing a matrix.
3. Although we like the discussion of the Leontief model, we think that the examples should discuss more carefully the translation of a verbal description of the economy to an input-output matrix. Too many of our students wound up with the transpose of the correct matrix, even after a careful in-class discussion of the problem formulation.
4. The discussion of Markov chains could be slightly expanded to include cyclic chains (so that ergodic chains can be classified). We also miss the discussion of absorption probabilities, and the related theorem, for absorbing chains; this material is included by Kemeny, Snell, and Thompson. Incidentally, we like to emphasize Markov chains in our courses because of the beautiful synthesis of ideas from probability theory and matrix algebra.

The text by Bittinger and Crown is attractively packaged and is written with the needs of the student reader in mind. The unusually wide pages include margins that contain statements of objectives for each section, marginal exercises carefully integrated with the text development, and space for working the marginal exercises. Where appropriate, the authors have even provided labeled coordinate axes to be used with the marginal exercises. These exercises, the odd-numbered problems, and the chapter tests have answers provided in an appendix. (More than a few of the answers turned out to be incorrect; we trust that these bugs and other misprints will disappear in subsequent printings.) The various learning aids provided were appreciated by our students and were cited in the students' favorable evaluation of the text.

The mathematical level of Bittinger and Crown is the lowest of the three texts, and the presentation is informal. Illustrations and loose descriptions often replace definitions and theorems. While we do not always insist on formality and rigor, we do cringe when we see misleading statements that substitute for definitions. Such is the case on page 317: "By a 'normal' class we mean that adding the test scores will not change the results." No mention appears of the normal density function, the normal approximation to a binomial, or assumptions that lead to a Central Limit Theorem result. The student could easily conclude that, after subtracting the mean and dividing by the standard deviation, *any* random variable is approximately normal.

In contrast, the introductions to matrix algebra and Gaussian elimination are superior in this text. Although a concise definition of a fully reduced matrix never appears, the development of Gaussian elimination is clear, complete, and well motivated. The development of regular Markov chains is equally commendable; however, absorbing chains are missing from that development.

In its first printing this text suffers from more misprints and errors than either of the others. When a mouse is "placed in a maze shown in an accompanying figure," the figure never appears (p. 339). Transition matrices are given whose row entries sum to more than one. Problems are posed that are ill defined (e.g., p. 282, no. 13); "answers" are nevertheless provided. It is unfortunate that a book with the pedagogical appeal of this one is marred by such blemishes.

In sum, we think that the text by Maki and Thompson is a viable alternative to the Kemeny, Snell, and Thompson classic, especially for those who want to emphasize discrete probability. The text of Bittinger and Crown is aimed at a different and less sophisticated audience. In a carefully revised and corrected edition, it could serve that audience very well.

JOHN D. EMERSON AND KARIN LARSON, Middlebury College

Accidental Nuclear War, Love Song, Auto Insurance. Three films produced by David Gillman. Each is 8 minutes in 16 mm sound and color. Available for rent or purchase from Pictura Films Distribution Corporation.

These three films are similar to the extent that they deal with some simple notion of probability, are technically well done, feature excellent acting, and have intelligent dialogue. The films are intended to be shown in an elementary undergraduate (or even high school) course in statistics.

The best of the three is *Accidental Nuclear War*. The action takes place in a general's office, where, in preparation for a press conference he will hold, the general is being briefed by a mathematician on the probability of an accidental nuclear war. Most of the mathematician's presentation is an explanation that $P(A \cap B) = P(A) \cdot P(B|A)$ and, hence, that the probability that there will never be a nuclear war is zero.

This film and *Love Song* were shown to a dozen mathematicians at the 1979 summer MAA meeting in Duluth. All agreed that the mathematics in *War* is well presented and understandable to the general audience. As an added bonus, the dialogue is humorous. Although these reviewers believe this to be a fine film, some viewers may find its satirization of the military objectionable.

Love Song is a cleverly done spoof-musical. Its story line revolves around a young woman's unsuccessful efforts to persuade a young man to have sexual relations with her. He refuses, because he fears he may have VD. His fears are not allayed by the negative result of a VD test, because the test is negative for 5 percent of those who *do* have VD. This sets the stage for the mathematical content of the film: an application of Bayes's Theorem. The probability that the male has VD, given that the VD test is negative, is not the same as the probability that the test is negative, given that he has VD (i.e., 5 percent). The film is confusing as to how and why the theorem applies in this situation. The physician attempts to explain with the aid of a chart that has Bayes's Theorem written on it. However, the chart is not large enough, nor does it stay in view long enough, to be of real assistance to the viewer. Instructors who intend to use this film should explain and illustrate Bayes's Theorem before showing the film.

In the opinion of these reviewers, *Love Song* is entertaining, but its mathematical content is minimal. We also feel compelled to mention that this film may be objectionable to many viewers: it has a sexual theme, it promotes male-female stereotyping, and it depicts the woman as having little concern for the possibility of her contracting VD.

Statistical dependence is the mathematical basis of *Auto Insurance*. In this film, a young black man talks with his insurance agent (an elderly white) about auto insurance rates because he has noticed that he and his black friends are paying more than the whites he knows for the same coverage. He is told that the insurance company does not discriminate on the basis of race; that insurance rates are determined by a computer, in accordance with certain formulas; and that the computer is not told the race of an applicant. However, an applicant's address is a factor in determining rates, and this may be indirect racial discrimination. Events A and B are *independent* if $P(A|B) = P(A)$, and it is known that accident rates are not independent of neighborhood; so neighborhood is used by the insurance company in calculating insurance risk. But the fact that two events are not independent does not mean that one causes the other—living in a black neighborhood may not make the man a poorer driver. Using a computer and formulas makes the process seem "mathematical" and impartial. But who gets to decide what factors should be considered? According to the black man in the film, "The people who decide this really decide everything, and they are white." This film is elementary and understandable. There is little real mathematics in it, but it forces one to consider some limitations and abuses of mathematical methods.

In summary, it is our opinion that instructors who wish to discuss social implications of mathematics and social issues, such as stereotyping, sexism, and racism, in their classes will find all three films of much value. For those who wish to discuss how statistics can be misused, *Accidental Nuclear War* and *Auto Insurance* are recommended.

JOSEPH A. GALLIAN, University of Minnesota, Duluth
DAVID S. WITTE, University of Chicago
JUDY LYNN SMITH, University of California, San Diego

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, S*(13-18), L***.** The William Lowell Putnam Mathematical Competition: Problems and Solutions: 1938-1964. A.M. Gleason, R.E. Greenwood, L.M. Kelly. MAA, 1980, xi + 652 pp, $35. [ISBN: 0-88385-428-7] Students and faculty will enjoy browsing, sampling, reading, and rereading the many gems in this outstanding collection. The solutions are enhanced by being set in a beautiful, inviting style. A thorough consideration of alternate solutions and a liberal inclusion of historical remarks and references to relevant literature combine to make the book especially useful as an instructional resource. A fitting and worthy tribute to the first twenty-five years of the Putnam Competition. LCL

General, S(14-17), P**, L**.** The Mathematical Experience. Philip J. Davis, Reuben Hersh. Birkhäuser, 1980, xix + 440 pp, $24. [ISBN: 3-7643-3018-X] A marvelous montage of history, reflection, commentary and philosophy that illuminates for the general reader the roots and importance of contemporary mathematics. This handsomely illustrated volume reveals as never before the mathematician's common experience of "guesswork, analogy, wishful thinking and frustration." It explores both "outer" and "inner" issues relating mathematis not just to science, but to other facets of man's experience, e.g., to religion, philosophy, pedagogy. Davis and Hersh's masterful style is enriched with an extraordinary sense of culture and history; the result is a true gem, one of the masterpieces of our age. LAS

General, S. Stamaza Matematike (Serbo-Croatian). (Along Mathematical Paths). Mirko Stojaković. Biblioteka Saznanja, University of Novi Sad, 1977, 253 pp. Forty-five popular, more or less mathematical, essays at various levels ranging from the trivial to the unusual. Twenty-five pages of portrait sketches and caricatures of mathematicians, by the author. RPB

General, P. Mathematics Curriculum Conference. Ed: Lynn Arthur Steen (Dept. of Math., St. Olaf College, Northfield, MN 55057), 1981, 91 pp, $4 (P). How can mathematics departments in small colleges integrate the new applied disciplines of computer science, statistics, and operations research with the traditional core subjects of algebra, topology, and analysis? This informal record of the conference proceedings held at St. Olaf College, November 14-15, 1980 based on transcripts and written notes, can serve to stimulate discussion. Included is the keynote address by William Lucas of Cornell University on "New Directions for Undergraduate Mathematics," two panel presentations, and discussions dealing with curricular issues for the 1980's. LCL

Precalculus, S(13). Assessing Competencies for Calculus: A Self-correcting Workbook. Raymond McGivney, James McKim, Benedict Pollina. Wadsworth, 1980, 48 pp, $4.95 (P). [ISBN: 0-534-00856-9] A supplement to a precalculus text Essential Precalculus (TR, January 1981). Problems are directly from calculus texts and might only confuse. LLK

Education, S(16), P. Problem-Solving Studies in Mathematics. Ed: John G. Harvey, Thomas R. Romberg. Wisconsin Res and Dev Center, 1980, 287 pp, $11 (P). Reports of nine interrelated studies originally done as doctoral dissertations. Investigation of mathematical problem solving in three areas; teaching heuristics, assessing problem solving performance, and establishing correlates with learner characteristics. Presents evidence that problem solving can be learned. Polya's influence is evident throughout. Includes extensive bibliography and review of recent (1969-78) research. MW

History, S*, P, L.** Sophie Germain: An Essay in the History of the Theory of Elasticity. Louis L. Bucciarelli, Nancy Dworsky. Stud. in History of Modern Sci., V. 6. Reidel, 1980, xi + 147 pp, $15.75 (P). [ISBN: 90-277-1135-6] Born in France in 1776, Sophie Germain won le prix extraordinaire of the French Academy in 1816 for her analysis of the newly discovered patterns of vibration of thin, flat elastic plates. This professional biography traces Germain's life--her struggle to be educated, her correspondence with contemporary (male) scientists--and her contributions to the mathematical theory of elasticity. LAS

Foundations, T*(15-16: 1), S, L*. Computability: An Introduction to Recursive Function Theory. Nigel Cutland. Cambridge U Pr, 1980, x + 251 pp, $47.50; $14.95 (P). [ISBN: 0-521-22384-9; 0-521-29465-7] Well-written introduction to recursion theory which assumes only familiarity with sets, functions and mathematical reasoning. First seven chapters cover basic theory, computable functions, decidable predicates, Church's thesis, the diagonal method, universal programs and functions, and recursive sets. Uses unlimited register machine to characterize computability. Final five chapters on more advanced topics--incompleteness, many-one and Turing reducibility, Kleene's two recursion theorems, and computational complexity. Includes exercises, numerous examples, suggestions for further study and comments on implications of results for computer programming. KS

Number Theory, S(18), P. Topics in Arithmetical Functions. J.-M. De Koninck, A. Ivić. Math. Stud., V. 43. North-Holland, 1980, xvii + 262 pp, $39 (P). [ISBN: 0-444-86049-5] A monograph which is primarily concerned with asymptotic formulae for the sums of reciprocals of non-negative arithmetic functions. Includes a number of the authors' hitherto unpublished results. CEC

Number Theory, S(18), P. p-adic Analysis: A Short Course on Recent Work. Neal Koblitz. London Math. Soc. Lect. Note Ser., No. 46. Cambridge U Pr, 1980, 163 pp, $14.95 (P). [ISBN: 0-521-28060-5] Discusses p-adic gamma, and log gamma functions, p-adic Gauss sums, the p-adic regulators of Leopoldt and Gross and their various connections. A sequel to the author's book in the Graduate Text Series (No. 58). Although there are no exercises, the book does contain many historical and motivational remarks and is very clearly written. SG

Linear Algebra, T(13-14: 1). Linear Algebra with Applications, Second Edition. Hugh G. Campbell. P-H, 1980, xii + 400 pp, $18.95. [ISBN: 0-13-536979-7] Exceptionally well integrated applications with lots of examples. Changes from the first edition (TR, November 1971; ER, June-July 1973) include more applications, early introduction to eigenvalues, and deletion of some abstraction. LLK

Algebra, T(16-17: 1, 2). Lehrbuch der Algebra: Unter Einschluss der linearen Algebra, Teil 1. Günter Scheja, Uwe Storch. B.G. Teubner, 1980, 408 pp, (P). [ISBN: 3-519-02203-6] First of three volumes which present a unified treatment of abstract and linear algebra. Volume 1 begins with review of induction, relations and cardinality, and then covers basic theory of monoids, groups, rings, modules, vector spaces, algebras and determinants. More advanced topics mentioned include affine spaces, chain conditions and exact sequences. Numerous exercises and examples. KS

Algebra, S(17-18), P. Darstellungstheorie von endlichen Gruppen. Wolfgang Müller. Teubner Stuttgart, 1980, ix + 211 pp, DM 24,80 (P). [ISBN: 3-519-02060-2] Ordinary and modular representation theory of finite groups. JD-B

Calculus, S(16-17). Analysis in mehreren Variablen. Theodor Bröcker. Teubner Stuttgart, 1980, vi + 361 pp, DM 29,80 (P). [ISBN: 3-519-02061-0] A concise, modern treatment of multivariate calculus. Includes chapters on differential equations and calculus on manifolds. Some problems. JD-B

Real Analysis, P*. Brownian Motion. T. Hida. Trans: T.P. Speed. Appl. of Math., No. 11. Springer-Verlag, 1980, xvi + 325 pp, $34. [ISBN: 0-387-90439-5] Rather than using the oft-discussed analysis of Brownian motion as a basic stochastic process, the author investigates the development of Wiener measure on function spaces. The treatment is heavily influenced by P. Levy's (1951) work and N. Wiener's 1923 paper and later work in cybernetics. An interesting treatment. TAV

Complex Analysis, T(18: 2), S, P. Riemann Surfaces. H.M. Farkas, I. Kra. Grad. Texts in Math., V. 71. Springer-Verlag, 1980, xi + 337 pp, $29. [ISBN: 0-387-90465-4] A surprisingly complete treatment covering topological, analytic and algebraic aspects of the theory. The style is very formal and a bit dry. The bibliography disappointingly short. TAV

Complex Analysis, T*(18: 1), S*, P. An Introduction to Nonharmonic Fourier Series. Robert M. Young. Pure and Appl. Math., No. 93. Acad Pr, 1980, x + 246 pp, $32. [ISBN: 0-12-772850-3] The first half of this aptly-titled book sets the stage for the main results on completeness and expansion properties of sets of complex exponentials in LP by presenting the relevant aspects of bases in Banach spaces and entire functions of exponential type. Suitable as a text for students with a background in complex analysis and some functional analysis. Well written. Many exercises. Good bibliography and notes. SES

Differential Equations, T(15-16), S. The Finite Element Method, A First Approach. Alan J. Davies. Clarendon Pr, 1980, xii + 287 pp, $49.50; $27.50 (P). [ISBN: 0-19-859630-8; 0-19-859631-6] An introduction for undergraduates, this text applies the finite element method to boundary and initial value problems described by partial differential equations. Developed from Poisson's equation, the method is then applied to time dependent and nonlinear problems, using both variational and weighted residual techniques. Requires linear algebra and vector calculus. Many exercises. Solutions. TRS

Differential Equations, T(16-18: 1, 2), S, P. The Finite Difference Method in Partial Differential Equations. A.R. Mitchell, D.F. Griffiths. Wiley, 1980, xii + 272 pp, $24.95. [ISBN: 0-471-27641-3] A major revision and updating of the 1969 Wiley volume Computational Methods in Partial Differential Equations (TR, March 1970). Clear exposition, but relatively few exercises. MU

Differential Equations, T*(15-16), S, L. State Models of Dynamic Systems: A Case Study Approach. N.H. McClamroch. Springer-Verlag, 1980, viii + 248 pp, $17.40. [ISBN: 0-387-90490-5] A text that grew out of the author's course for undergraduate students of electrical or computer engineering at Michigan. Case studies, based primarily on ordinary differential equations, expose the important system concepts, e.g., model, time invariance, approximation, simulation, feedback, stability. Presumes solid background in ordinary differential equations and elementary mechanics and circuits. Many modeling exercises. TRS

Differential Equations, P. Differential Equations with Small Parameters and Relaxation Oscillations. E.F. Mishchenko, N. Kh. Rozov. Trans: F.M.C. Goodspeed. Plenum Pr, 1980, x + 228 pp, $29.50. [ISBN: 0-306-39253-4] Mainly on results of Pontryagin and the authors on asymptotic and almost-discontinuous periodic solutions of second order systems of ordinary differential equations in which small parameters multiply certain derivatives. Sixty-six references to well-known books,

reports of symposia and papers, including eighteen with Pontryagin, Mishchenko and Rozov as author or co-author. JK

Differential Equations, P. Introduction à la théorie des équations aux dérivées partielles linéaires. Jacques Chazarain, Alain Piriou. Gauthier-Villars, 1981, 466 pp, (P). [ISBN: 2-04-012157-9] An introduction to the modern theory of partial differential equations. Topics: distribution theory, Sobolev spaces, stationary phase approximation of oscillatory integrals, pseudo-differential operators, elliptic problems, evolution equations, hyperbolic problems, microlocalization. TRS

Differential Equations, T(17-18: 2), S, L. Dynamical Systems and Evolution Equations, Theory and Applications. J.A. Walker. Math. Concepts and Methods in Sci. and Eng., V. 20. Plenum Pr, 1980, viii + 236 pp, $29.50. [ISBN: 0-306-40362-5] Addresses the study of physical systems whose evolution in time can be described by differential equations. Starting with systems modeled in finite dimensions, the author develops a metric and Banach space framework for abstract evolution equations and describes the tools from topological dynamics (e.g., Liapunov functions) that are basic to the study. Exercises. TRS

Numerical Analysis, S(18), P. Numerical Solution of Systems of Nonlinear Equations. J.C.P. Bus. Math. Centre Tracts, No. 122. Math Centrum, 1980, vi + 265 pp, Dfl. 32 (P). [ISBN: 90-6196-195-5] A monograph devoted to Newton-like methods, giving comprehensive analysis of convergence, synthesis, and experimental evaluations of the methods. Includes comparison with Brown's method, which is competitive with those of Newton type. TRS

Functional Analysis, T(17), S. Introduction to H_p Spaces, With an Appendix on Wolff's Proof of the Corona Theorem. Paul Koosis. London Math. Soc. Lect. Note Ser., No. 40. Cambridge U Pr, 1980, xv + 376 pp, $19.95 (P). [ISBN: 0-521-23159-0] In keeping with lecture notes, the style is informal and relaxed. The aim is to present the concrete aspects of H_p theory in the simplest cases. A very readable treatment requiring only introductory courses in real and complex variables and a little functional analysis. Wolff's proof mentioned in the title is a very nice application of the theory of bounded mean operators. TAV

Functional Analysis, T*(16-18: 1, 2), P, L. Elements of Applicable Functional Analysis. Charles W. Groetsch. Pure and Appl. Math., V. 55. Dekker, 1980, x + 300 pp, $24.75. [ISBN: 0-8247-6986-4] An excellent introductory survey of Banach and Hilbert spaces. Faithful to its title, the text includes many applications, e.g., quadrature rules, harmonic analysis, interpolation, Kuhn-Tucker theorem, Lasalle's bang-bang theorem in control theory, Lax-Milgram lemma and boundary value problems, Ritz-Galerkin method, Sard's generalized splines. Requires background in linear algebra and rigorous analysis but little Lebesgue theory. Many exercises. TRS

Functional Analysis, P. Lecture Notes in Mathematics-823: Integral Operators in Potential Theory. Josef Král. Springer-Verlag, 1980, 171 pp, $11.80 (P). [ISBN: 0-387-10227-2] Single and double layer integral potentials are developed and applied to the Neumann and Dirichlet problems, respectively. The normal derivatives of these operators are then investigated. Contains a very extensive bibliography. TAV

Functional Analysis, S*(16-18), P*, L*. Studies in Functional Analysis. Ed: R.G. Bartle. Stud. in Math., V. 21. MAA, 1980, xi + 227 pp, $19. [ISBN: 0-88385-121-0] Five fascinating expository articles on recent developments in functional analysis: Numerical Ranges by Bonsall and Duncan; Projection Operators in Approximation Theory by Cheney; Complementably Universal Separable Banach Spaces: An Application of Counterexamples to the Approximation Problem by W.B. Johnson; Integral Representations for Elements in a Convex Set by Phelps; and Aspects of Banach Lattices by Schaefer. TRS

Functional Analysis, P. Lecture Notes in Mathematics-793: A Groupoid Approach to C*-Algebras. Jean Renault. Springer-Verlag, 1980, 160 pp, $11.80 (P). [ISBN: 0-387-09977-8] Modeling the construction of the C*-algebra of a transformation group, the author constructs and studies the C*-algebra of a locally compact groupoid with a fixed Haar system. TRS

Functional Analysis, T*(16), L. Applied Functional Analysis. Jean-Pierre Aubin. Trans: Carole Labrousse. Wiley, 1979, xv + 423 pp, $26.95. [ISBN: 0-471-02149-0] This book presents the important results of linear, convex, and nonconvex functional analysis in the framework of Hilbert spaces, studying many applications to optimization theory, game theory, equilibrium theory in economics, numerical analysis, systems theory, approximation theory, optimal control, distributions and Sobolev spaces, and boundary value problems for elliptic and parabolic partial differential equations. Restriction to Hilbert space allows simple development of theory as well as serious consideration of applications. Presumes linear algebra and metric topology. Few exercises. TRS

Optimization, T(15-17: 2), L?. Optimization: Theory and Applications. S.S. Rao. Wiley Eastern, 1979, xiv + 711 pp, $19.95. [ISBN: 0-85226-756-8] Comprehensive survey of optimization techniques, oriented toward engineering applications. Linear and non-linear programming are treated thoroughly. Topics covered also include geometric, dynamic, stochastic and integer programming. References and exercises for each chapter. JRG

Optimization, T(16-17: 2). Introduction to Operations Research, Third Edition. Frederick S. Hillier, Gerald J. Lieberman. Holden-Day, 1980, xiv + 829 pp, $28.95. [ISBN: 0-8162-3867-7] Third edition of a standard text. Deletes some of the mathematically demanding sections (but still addresses

engineers, others with mathematical background), adds more practical material, up-dates treatment of integer programming. (First Edition, TR, February 1968; Second Edition, TR, June-July 1976.) AWR

Optimization, P. Approaches to the Theory of Optimization. J. Ponstein. Tracts in Math., No. 77. Cambridge U Pr, 1980, xii + 205 pp, $36.50. [ISBN: 0-521-23155-8] The "Approaches" of the title emphasize topological aspects of optimization; the primal-dual nature of optimization problems is thoroughly discussed. Frequent observations on the nature of the subject matter, as well as the theoretical results. Engaging. JRG

Analysis, S(17-18), P. Analysis, Part II: Integration, Distributions, Holomorphic Functions, Tensor and Harmonic Analysis. Krzysztof Maurin. Reidel, 1980, xvii + 829 pp, $76.30. [ISBN: 90-277-0865-7] Starting with general topology and general integration theory (including the integrals of Radon, Wiener and Bochner), this encyclopedic text develops surface integrals and their applications, complex functions of one and several variables, partitions of unity, distributions, and harmonic analysis. No exercises. TRS

Analysis, T(16-18: 1, 2), S, L. Mathematical Foundations in Engineering and Science: Algebra and Analysis. Anthony N. Michel, Charles J. Herget. P-H, 1981, xi + 484 pp, $25.95. [ISBN: 0-13-561035-4] A unified overview of algebra and analysis designed to provide beginning graduate students in engineering and science with appropriate mathematical background. Begins with a relatively brief chapter on introductory set theory, followed by one on abstract algebra, two on linear algebra, one on metric spaces and two on functional analysis. No applications to engineering or science appear in the text, and there is not much breathing room for the reader. Exercises, many asking for proofs, are woven into the text. The abstract nature of the material will provide rough mathematical going for the average engineering or science graduate student. Small number of references, mostly to standard and well-known books are given at chapter ends. JK

Analysis, P. Harmonic Analysis and Representations of Semisimple Lie Groups. Ed: J.A. Wolf, M. Cahen, M. De Wilde. Math. Physics & Appl. Math., V. 5. D. Reidel, 1980, viii + 495 pp, $66. [ISBN: 90-277-1042-2] Lectures delivered at the NATO Advanced Study Institute at Liege, Belgium in September 1977 on "Representations of Lie groups and harmonic analysis." Starting with a general background on the representation of locally compact groups and semisimple Lie groups, the book proceeds to harmonic analysis for reductive groups and infinitesimal methods in representation theory and beyond. Note the price. SES

Differential Geometry, S(18), P. Lecture Notes in Mathematics-812: Toroidal Compactification of Siegel Spaces. Yukihiko Namikawa. Springer-Verlag, 1980, viii + 162 pp, $11.80 (P). [ISBN: 0-387-10021-0] If D is a hermitian bounded symmetric domain and I is a discrete subgroup of the Lie group of biholomorphic automorphisms of D then I/D is a normal complex analytic space. Discussion here centers on compactification of I/D, particularly recent work by Satake and Mumford. Bibliography, indexes. JS

Topology, P. Two-Bridge Knots Have Property P. Moto-o Takahashi. Memoirs No. 239. AMS, 1981, iii + 104 pp, $6 (P). [ISBN: 0-8218-2239-X] A knot has property P if every 3-manifold obtained by a non-trivial surgery along the knot fails to be simply connected. JAS

Statistics, S(14-18), L. Biostatistics Casebook. Ed: Rupert G. Miller, Jr., et al. Wiley, 1980, xii + 238 pp, $14.95 (P). [ISBN: 0-471-06258-8] Case histories of twelve biostatistical applications mostly from the Stanford Medical School. FLW

Statistics, T(16-18: 1), S, P, L. Robust Estimation. Robert G. Staudte, Jr. Pure and Appl. Math., No. 53. Queen's U, 1980, iv + 111 pp, (P). An introduction to the topic, "especially that portion which can be nicely described in the language of functionals on the class of probability models." FLW

Statistics, T(15-16: 1, 2), S. Applied Statistical Techniques. K.D.C. Stoodley, T. Lewis, C.L.S. Stainton. Ellis Horwood, 1980, 310 pp, $56.95. [ISBN: 0-85312-157-5] A brief review of basic statistics followed by relatively non-mathematical discussions of regression analysis, design of experiments, nonparametric methods, sampling inspection, sample surveys, and forecasting. Note the exorbitant price. FLW

Statistics, T(16-18: 1), S, P, L. Incomplete Block Designs. Peter W.M. John. Lect. Notes in Stat., V. 1. Dekker, 1980, vii + 101 pp, $17.50 (P). [ISBN: 0-8247-6995-3] An introduction to the topic. No exercises. No index. FLW

Statistics, P. Statistical Analysis of Weather Modification Experiments. Ed: Edward J. Wegman, Douglas J. DePriest. Lect. Notes in Stat., V. 3. Dekker, 1980, x + 145 pp, $24.75 (P). [ISBN: 0-8247-1177-7] Widely differing answers to two fundamental questions: Can weather be effectively modified? If so, can these modifications be controlled? Papers based on a workshop called to discuss the Santa Barbara Convective Seeding Test Program, 1967-1971. LAS

Statistics, P. Hypothesis Testing with Complex Distributions. Kenneth S. Miller. Krieger, 1980, viii + 176 pp, $14.50. [ISBN: 0-88275-463-7] Investigation of complex probability distributions, emphasizing derivation of closed forms of various density functions. LAS

Statistics, S(14-15). Biomathematik für Mediziner. Edward Walter. Teubner Stuttgart, 1980, 206 pp,

DM 19,80 (P). [ISBN: 3-519-12049-6] A survey of definitions and results from combinatorics, calculus, statistics (most of the book), and data processing for medical students. This is a rewritten version of the 1974 edition (TR, October 1975) in response to the 1978 standards for the German medical profession. JAS

Computer Programming, T*(13-18: 1, 2), S. PL/I Structured Programming, Second Edition. Joan K. Hughes. Wiley, 1979, xiii + 825 pp, $18.95. [ISBN: 0-471-01908-9] Beginning with chapter one the reader can code, compile, and execute complete PL/I programs. Techniques of structured programming are stressed from the start. Helpful features of each chapter include sections with checkpoint questions, practice problems, debugging techniques, and case studies. Appendix. Answers to checkpoint questions. Glossary. Index. RJA

Computer Programming, T, S, P. A Guide to PL/I and Structured Programming, 3rd Edition. Seymour V. Pollack, Theodor D. Sterling. HR&W, 1980, xxvi + 646 pp, $18.95. [ISBN: 0-03-55821-5] A comprehensive, thorough and readable introduction to PL/I. Includes many examples and exercises. In this edition the principles of structured programming are emphasized more fully. CEC

Computer Programming, T(13-14: 1). Pascal Programming Structures, An Introduction to Systematic Programming. George W. Cherry. Reston, 1980, xiii + 314 pp, $12.95 (P). [ISBN: 0-8359-5463-3] A well-conceived introductory text with ample explanations of programming concepts and program examples. Some students may find it too wordy. Includes exercises. JL

Computer Programming. Problems for Computer Solutions Using FORTRAN. Henry M. Walker. Winthrop, 1980, x + 203 pp, $12.95 (P). [ISBN: 0-87626-654-5] A good source of computer problems. The book is in two parts, an introduction to Fortran programming and a collection of programming problems oriented to many interests (i.e., calculus, physics, chemistry, linear algebra). LLK

Computer Science, T(15-18: 1, 2), S, P, L. Distributed Micro/Minicomputer Systems: Structure, Implementation, and Application. Cay Weitzman. P-H, 1980, xiii + 403 pp, $22.50. [ISBN: 0-13-216481-7] Begins with an overview of multimicro- and multiminicomputer systems. Includes multiminicomputer architecture, multimicro- and minicomputer software, off-the-shelf multiminicomputer hardware, design based on process characterization, examples of applications (Bank of America, Ford's LNA, NASA, signal processing), future trends, and an appendix on different data link control protocols. Chapter problems. Sample problem solutions. Index. RJA

Computer Science, T*(13-18: 1, 2), S, P, L. Computer Programming and Architecture: The VAX-11. Henry M. Levy, Richard H. Eckhouse, Jr. Digital Pr, 1980, xxi + 407 pp, $15. [ISBN: 0-932376-07-X] Text has two parts: the computer's architecture from the perspective of an assembly language programmer, and the aspects of the architecture used by the operating system and the strategies used to manage hardware resources. Written in a general manner using the VAX as a specific example. Each chapter has a summary, references, and exercises. Appendixes. Bibliography. Index. RJA

Computer Science, S(15-18), P, L. Advances in Computers, V. 19. Ed: Marshall C. Yovits. Acad Pr, 1980, x + 351 pp, $36.50. [ISBN: 0-12-012119-0] Contains five survey articles on hardware of data base computers, structure of parallel algorithms, software in numerical analysis, data clustering, and the sociology of computing. References at end of individual articles. Author index. Subject index. List of the contents of previous volumes. RJA

Computer Science, P. A General Theory of Optimal Algorithms. J.F. Traub, H. Woźniakowski. Acad Pr, 1980, xiv + 341 pp, $36. [ISBN: 0-12-697650-3] Are some problems intrinsically harder than others? Problem complexity is defined as the complexity of the optimal algorithm for solving the problem. The authors present here a foundation and general framework for the study of optimal algorithms for problems that are solved approximately. This highly technical monograph constitutes a report on work in progress; it includes numerous conjectures, open problems, and alternative models which need to be explored. LCL

Computer Science, P. Handbook of Computer-Aided Composition. Arthur H. Phillips. Dekker, xviii + 434 pp, $55. [ISBN: 0-8247-6963-5] Survey of issues and hardware concerning the use of computers in composition, including discussion of fonts, keyboards, OCR, word processing, and photocompositors. One chapter is devoted to special problems of scientific and mathematical composition. Of more historic than current value, it nevertheless sheds valuable insight on the work currently going on in computerization of mathematical text. LAS

Computer Science, S*, P*. Z80 Assembly Language Programming. Lance A. Leventhal. Osborne/McGraw-Hill, 1979, xx + 609 pp, $16.99 (P). [ISBN: 0-931988-21-7] An excellent presentation of Z80 assembler language using hundreds of sample routines and exercises. The instructions themselves are well presented, each with an explanation of "who cares." There are tables ordered by every conceivable way of looking for a command (alphabetic, numerical, frequency of use, computer architecture). Includes considerable exposition of why and how program samples work via input and output samples and "what if" discussions. JAS

Systems Theory, S(14-18), P, L. Fuzzy Sets: Theory and Applications to Policy, Analysis and Information Systems. Ed: Paul P. Wang, S.K. Chang. Plenum Pr, 1980, ix + 413 pp, $42.50. [ISBN: 0-306-40557-1] Nine papers on recent theoretical developments and eighteen papers on applications of fuzzy sets. Included are applications to statistics, pattern recognition, logic, cost-benefit analysis,

small group dynamics, interpersonal communication, reliability theory, and systems. FLW

Applications (Actuarial Science), P. Computational Probability. Ed: P.M. Kahn. Acad Pr, 1980, xi + 340 pp, $21. [ISBN: 0-12-394680-8] Collection of papers presented at the Actuarial Research Conference on Computational Probability and related topics held at Brown University in August, 1975. Topics covered include computational probability, computational statistics, computational risk theory, analysis of algorithms, numerical methods, and notation and computation. RSK

Applications (Biology), S(16-18), P. Modeling and Differential Equations in Biology. Ed: T.A. Burton. Lect. Notes in Pure and Appl. Math., V. 58. Dekker, 1980, viii + 277 pp, $35 (P). [ISBN: 0-8247-1075-4] The proceedings of a CBMS Conference at Southern Illinois University, June 5-9, 1978. Chapters describe how stability theory of differential equations is used in modeling microbial competition, predator-prey systems, humoral immune response, and dose and cell-cycle effects in radiotherapy. TRS

Applications (Control Theory), P. Lecture Notes in Control and Information Sciences-27: Extensions of Linear-Quadratic Control Theory. D.H. Jacobson, et al. Springer-Verlag, 1980, xi + 288 pp, $19.20 (P). [ISBN: 0-387-10069-5] These notes provide a readable and reasonably complete survey of the state-of-the-art in multivariable linear-quadratic optimal control and estimation. Each of the fifteen lectures presented contains a useful bibliography. TAV

Applications (Engineering), P. Discrete Linear Control: The Polynomial Equation Approach. Vladimír Kucera. Wiley, 1979, 206 pp, $39.75. [ISBN: 0-471-99726-9] The claim that the reader needs only an elementary knowledge of polynomials and matrices should be ignored. A background in traditional methods (complex variables) is necessary to help the reader understand what problem is being addressed. AWR

Applications (Game Theory), T(16-17: 1), S, P, L*. Game Theory: Mathematical Models of Conflict. A.J. Jones. Halsted Pr, 1980, 309 pp, $45. [ISBN: 0-470-26870-0] A sophisticated, self-contained introduction to game theory which includes a general introduction and chapters on non-cooperative games, linear programming and matrix games, cooperative games and bargaining models. Includes lots of good problems along with detailed solutions in the back. Assumes linear algebra, calculus and some topological ideas which can be done intuitively. CEC

Applications (Physics), P?. Lie Groups: History, Frontiers and Applications, V. X: Quantum Statistical Mechanics and Lie Group Harmonic Analysis, Part A. Norman Hurt, Robert Hermann. Math Sci Pr, viii + 251 pp, $30. [ISBN: 0-915692-30-9] A nonrigorous exposition of mathematical statistical mechanics with an emphasis on differential geometry and Lie group theory. AO

Applications (Psychology), P. The Interpretation of Visual Motion. Shimon Ullman. MIT Pr, 1979, 229 pp, $17.50. [ISBN: 0-262-21007-X] A study of visual perception using ideas from artificial intelligence. Two major problems are studied: the correspondence problem (when do two images represent the same object at different times?) and the three-dimensional interpretation of a changing image once a correspondence has been established. JAS

Applications (Simulation), P. Computer Simulation 1951-1976: An Index to the Literature. Per Holst. Mansell, 1979, xiv + 438 pp. [ISBN: 0-7201-0734-2] Contains two types of indexes: a multipermuted title index and an author index. The entries represent innovative, new, or summary works on simulation technology. Included are entries that preserve the tricks of the trade, present model data, and results presenting a unique contribution to the field. List of sources for the entries. RJA

Applications (Social Science), T(16-18: 1), S, P, L. Introduction to Mathematical Consensus Theory. K.H. Kim, Fred W. Roush. Lect. Notes in Pure and Appl. Math., V. 59. Dekker, 1980, vi + 179 pp, $25 (P). [ISBN: 0-8247-1001-0] Theory of social welfare functions (e.g., Arrow's theorem; approval voting) and related aspects of game theory (e.g., Shapley value), from basic structure (binary relations, Boolean matrices) to current research. Austere definition-theorem-proof style, with only occasional remarks and commentary. LAS

Applications (Social Science), P, L. Mathematical Models as a Tool for the Social Sciences. Ed: Bruce J. West. Gordon, 1980, 120 pp, $26.50. [ISBN: 0-677-10390-5] Papers on contracts, memory, historiography, speculation and economics representing a variety of approaches to mathematical modelling: idealization, refinement, axiomatic, stochastic, normative. An outgrowth of an interdisciplinary seminar at the University of Rochester. Typescript text. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; RPB: Ralph P. Boas, Northwestern Univ. MB: Murray Braden, Macalester; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

At the California Institute of Technology Professor *W.A. J. Luxemburg* has received a Senior U. S. Scientist Award from the Humboldt Foundation of West Germany for his pioneering work in non-standard analysis. *W. Hugh Woodin* has been appointed Assistant Professor.

California State University at Sacramento: Assistant Professor *Elaine Alexander* has been promoted to Associate Professor. *H. Stewart Moredock* retired June 30, 1980, with the title Professor Emeritus of Mathematics.

University of California, Berkeley: *Jerrold E. Marsden* and *Alan Weinstein* will be Miller Research Professors during 1981-82.

Carleton College: Professor *John Dyer-Bennett* retired in June 1980 with emeritus rank. *Stan Seltzer* has been appointed Assistant Professor.

Glassboro State College: Associate Professor *Francis E. Masat* has been promoted to Professor and Assistant Professor *Robert D. Mitchell* has been promoted to Associate Professor. *Patrick L. Smiley* retired June 30, 1980, with emeritus rank.

Miami University (Oxford, Ohio): *Thomas Grilliot* has accepted an appointment as Lecturer. Assistant Professor *Clyde D. Hardin,* Jr., has resigned to join the faculty of the University of Wisconsin, Milwaukee. Associate Professor *Donald O. Koehler* has been promoted to Professor and Assistant Professor *Robert M. Dieffenbach* has been promoted to Associate Professor.

Oregon State University: Professor *James R. Brown* has resigned to be a fulltime law student. Associate Professor *Bent E. Petersen* has been promoted to Professor. *Harry E. Goheen* has retired, effective December 31, 1980, with the title Professor Emeritus.

University of Washington: Associate Professor *Thomas Hungerford* has resigned to become Professor and Department Chairman at Cleveland State University. Assistant Professor *Gerald B. Folland* has been promoted to Associate Professor.

Western Michigan University: Professor *Yousef Alavi* is the first recipient of the University's Distinguished Service Award. Professor *Gary Chartrand* received the University's Distinguished Faculty Scholar Award for 1979-80. Associate Professor *Joseph Buckley* has been promoted to Professor.

Westminster College: *Ben Andrew Budde* has been appointed Assistant Professor. Professor *Clyde Raymond Barrow* retired June 1, 1980, with the title Professor Emeritus.

Edgar A. Franz, Professor of Mathematics and Chairman of the Department at Illinois College, was installed as the Hitchcock Professor of Mathematics at a convocation January 19, 1981.

Morris Morduchow, Professor of Mechanical and Aerospace Engineering at the Polytechnic Institute of New York, has received the 1980 I.B. Laskowitz Award of the New York Academy of Sciences for his research contributions in aerospace engineering. The award consists of a gold medal and citation.

Jerry P. King, Professor of Mathematics and Associate Dean of Arts and Science at Lehigh University, was named Dean of the Graduate School effective January 1, 1981.

James Stormer, formerly a visiting lecturer at Holy Cross College, has been appointed Assistant Professor at the U.S. Naval Academy.

Gerald A. Heuer of Concordia College is on sabbatical leave for the academic year 1980-81. He is serving as Visiting Professor at Washington State University.

Nachman Aronszajn, Adjunct Professor of Mathematics at Oregon State University, died Feb. 5, 1980. He had been a member of the Association for twenty-seven years.

John S. Biggerstaff of Portland, Oregon, died at the age of eighty-four. He was a member of the Association for fifty-seven years.

Sister *Mary Michael Maloney* died Jan. 2, 1981, at the age of ninety. She was a member of the Association for thirty-five years.

Alan H. Marshall of Los Alamos died October 20, 1980, at the age of forty-five. He was a member of the Association for one year.

James D. McKnight, Jr., of the University of Miami, died Jan. 15, 1981, at the age of fifty-two. He was a member of the Association for twenty-seven years.

Edgar R. Mullins, Jr., Director of Computer Education at Swarthmore College, died Jan. 3, 1980, at the age of sixty-six. He was a member of the Association for twenty-seven years.

Henry Allen Robinson, retired head of the Department of Mathematics at Agnes Scott College, died Jan. 7, 1981, at the age of seventy-nine. A member of MAA for nearly sixty years, he was one of the organizers of the Southeastern Section in 1922. He was secretary of that section from 1933 to 1959 and was a governor of the Association.

SIAM NATIONAL MEETING

The Society for Industrial and Applied Mathematics will meet June 8-10, 1981, at Rensselaer Polytechnic Institute, Troy, N.Y. Themes of the meeting are mathematical aspects of combustion, scientific and statistical computing, and algebraic and discrete methods. *Garrett Birkhoff* will deliver the 1981 John Von Neumann Lecture. Another featured speaker will be *Cheryl Griffiths Tsopf*, the current Congressional Science Fellow in Mathematics.

STONE RINGS TOUR

Lyman C. Peck is conducting a tour of the stone rings of Scotland and England, July 1-17, 1981. Many of the 900 rings in existence appear to give evidence of surprising mathematical insights, as discussed in THE AMERICAN SCIENTIST, January 1972. For tour information write or call

Professor Lyman C. Peck
Department of Mathematics & Statistics
Bachelor Hall
Miami University
Oxford, Ohio 45056 Tel. (513) 529-5818

COMBINATORIAL PROBLEM-SOLVING SHORT COURSE

The Northeastern Section of MAA is sponsoring a short course in Combinatorial Problem-Solving at the University of Maine, Orono, June 15-19, 1981. The principal lecturer will be *Alan Tucker*, Chairman of the Department of Applied Mathematics and Statistics at SUNY, Stony Brook.

The course will examine the mathematical reasoning underlying combinatorial problem-solving in the mathematical sciences, with the goal of preparing mathematics faculty to teach a course in applied combinatorial mathematics. Course content will include theories and heuristic approaches to basic problems in enumeration and graph theory, with applications to computer science.

For further information, write:

Don Small, Colby College, Waterville, Maine 04901
Gratten Murphy, Department of Mathematics
University of Maine, Orono, Maine 04469

1982 INTERNATIONAL CONFERENCE ON TEACHING STATISTICS

The International Statistical Institute has announced a conference to be held in Sheffield, England, August 8-13, 1982. The objective of the Conference is to improve the quality of statistics teaching on a world-wide basis. Key goals include fostering international co-operation among teachers of statistics and promoting the interchange of ideas about teaching materials, methods, and content. Teaching from the school to the college level will be included.

For a copy of the first announcement, write to:

The Conference Secretary
International Conference on the Teaching Statistics
Department of Probability and Statistics
The University
Sheffield S3 7 RH
England

CONFERENCE AT MIAMI (OHIO)

The Ninth Annual Mathematics and Statistics Conference at Miami University, Oxford, Ohio, will be held September 25-26, 1981. The theme for this year's conference will be "Emerging Trends in Mathematics and Its Instruction." Featured speakers will include Paul R. Halmos, Indiana University; Alan Tucker, SUNY at Stony Brook; and James W. Wilson, University of Georgia. *Contributed papers relating to the general theme and appropriate for a general audience of mathematicians are welcome. Of particular interest will be papers dealing with emerging trends in a particular area of mathematics or mathematics education.* Abstracts should be sent by June 1, 1981, to David Kullman or Lyman Peck, Department of Mathematics and Statistics, Miami University, Oxford, Ohio, 45056. Information concerning preregistration, housing, etc., will be available from this address after July 15.

AMERICAN SOCIETY FOR CYBERNETICS TO MEET

The American Society for Cybernetics will meet October 31 to November 2, 1981, at the Washington Hilton Hotel, Washington, D.C., immediately following a meeting of the American Society for Information Science at the same hotel. The theme of the meeting will be "The New Cybernetics." One of the objectives of the meeting is to redefine the field of cybernetics and to provide a focus for the research efforts of the rejuvenated society and its membership.

Information may be obtained from?

Dr. Laurence D. Richards
Department of Administrative Science
Colby College
Waterville, Maine 94901

CRYPTOLOGIA UNDERGRADUATE PAPER COMPETITION

CRYPTOLOGIA, a journal devoted to all aspects of cryptology, announces a prize of three hundred dollars to be awarded to an undergraduate for the best paper submitted by Jan. 1, 1982. The competition is being underwritten by a gift from Professor *Boshra H. Makas* of the Department of Mathematics, St. Peter's College, Jersey City, N.J.

For the rules of the competition and other information, write to CRYPTOLOGIA, Albion College, Albion, Michigan 49224

THE MATHEMATICAL ASSOCIATION OF AMERICA
THE SIXTY-FOURTH ANNUAL MEETING OF THE ASSOCIATION

The Sixty-Fourth Annual Meeting of the Mathematical Association of America was held at the San Francisco Hilton in San Francisco, California, from Wednesday through Sunday, January 7-11, 1981, in conjunction with meetings of the American Mathematical Society, the Association for Symbolic Logic, the Association for Women in Mathematics, the Conference Board of the Mathematical Sciences, the National Council of Teachers of Mathematics, and the Mathematicians Action Group. Sessions of the MAA were held on Friday and Saturday mornings and on Sunday morning and afternoon. There were 3165 registrants, including 1635 members of the Association.

Presiders at the lectures were MAA President Dorothy L. Bernstein and Professors Lenore Blum, Gulbank D. Chakerian, William G. Chinn, Hugh M. Edgar, Hans Samelson, Gerald L. Alexanderson, Jean Pedersen, and David Gale.

FIRST SESSION OF THE ASSOCIATION (JOINT WITH NCTM)

"The Differing Ideals of Dedekind and Kronecker" by Professor Harold M. Edwards, NYU.

Richard Dedekind and Leopold Kronecker were the two pioneers in generalizing Kummer's theory of "ideal prime factors" from cyclotomic fields to arbitrary algebraic number fields. Coincidentally, they were also two of the most influential thinkers of their time on issues concerning the foundations of mathematics. Their views on the foundations differed sharply and, as a result, their formulations of the theory of "ideals" differed too. These differences were examined.

"Agenda for Action": Progress and Problems", by Professor Max A. Sobel, NCTM President.

The National Council Teachers of Mathematics has conducted an extensive survey of the opinions of many sectors of society concerning the directions that school mathematics programs should be taking in the 1980s. This survey, funded by the National Science Foundation, was called "Priorities in School Mathematics" (PRISM). Using this base, information from the mathematics assessments of the National Assessment of Educational Progress, and the concentrated efforts of professional educators, the NCTM has prepared a document entitled AN AGENDA FOR ACTION with their recommendatons for school mathematics of the 1980s. These were reviewed together with a discussion of progress and problems relative to their implementation.

"Contact Measures in Integral Geometry", by Professor William J. Firey, Oregon State University.

For compact convex sets K, K' with boundaries gK, gK' in a Euclidean space contact measures are motion-invariant measures of sets of motions f such that fK meets K' but not its interior. These are defined for subsets S, S' of gK, gK'. When S, S' are inverse spherical images of Borel sets of directions, the measure is bilinear in the area functions of Fenchel and Jessen. When S, S' are intersections of gK, gK' with spatial Borel sets, the measure is bilinear in Federer's curvature measures.

"Applications from UMAP", by Ross L. Finney, UMAP Director.

SECOND SESSION OF THE ASSOCIATION (JOINT WITH NCTM)

Panel Discussion: Panel on Gifted Students.

A panel discussion moderated by Professor Jean J. Pedersen, University of Santa Clara, with the presentations "Local Programs for Gifted High School Students" by Professor Pedersen, "Student Science Training Projects" by Professor Edmund J. Deaton, San Diego State University, "Problem Competitions" by Mr. Lyle Fisher, Redwood High School, "Project MEGSSS ("Mathematics Education for Gifted Secondary School Students)" by Mr. Joel Schneider, CEMREL, Inc.

Professor Deaton said that Student Science Training Programs have been supported by NSF since 1958. Approximately 115 grants are awarded each year to colleges or non-profit educational organizations. SSTP's offer opportunities to outstanding high school students to spend several weeks during the summer, or throughout the academic year, in specially designed courses or in research. Most programs are residential, for 35-40 students between their junior and senior years. Participants are expected to pay as much of the room and board costs as possible. Many programs focus on the under-represented in science or on students with limited science educational opportunities. A new focus now calls for programs to be designed for the junior high school student.

Mr. Schneider said that CEMREL established its MEGSSS Center in September 1978 in St. Louis. This center provides a special curriculum, the Elements of Mathematics, which is designed for students superior in mathematics and reasoning ability. The program is a complete mathematics curriculum taking a unique approach to mathematics. Not only are traditional topics treated rigorously

from an advanced point of view, but much of the content of a strong undergraduate major in mathematics is included for those who complete the program. The project indentifies students by grade seven, enables them to remain within their geographic, social, and chronological environment, and provides an economical means to provide the program by drawing from a large population base.

"How the Mathematical Sciences Section at the National Science Foundation Works", by Dr. William G. Rosen, Mathematical Sciences Section Head, NSF.

"The New Soviet Challenge in Mathematics and Science Education and Manpower Training," by Professor Izaak Wirszup, University of Chicago.

Professor Wirszup examined the recent Soviet educational mobilization, emphasizing a multi-track secondary school system comprising general education schools, specialized professional establishments (tekhnikums), and technical-vocational schools. He described the curriculum reforms in mathematics and the sciences, including the role of the USSR Academy of Sciences and the research conducted at the USSR Academy of Pedagogical Sciences in the psychology and methods of learning and teaching. Programs for the gifted and entrance examinations to institutions of higher learning were cited, and some comparisons with the United States' school system and manpower training programs were drawn.

THIRD SESSION OF THE ASSOCIATION

"Curvature", by Professor Robert Osserman, Stanford University.

Professor Osserman demonstrated that the concept of curvature is a central one in differential geometry. There are certain intuitive notions about what curvature should represent, and they can be seen underlying the many different quantities that are labeled "curvature" in one context or another. The history of the curvature concept was traced through from the simplest cases to see how one is led naturally to such expressions as the generalized Gauss-Bonnet integrand on Riemannian manifolds.

"Patterns of Problem Solving", Professor Moshe F. Rubinstein, Engineering Systems, UCLA.

"Patterns of Problem Solving" included the following main topics: Anatomy of a Problem; Concepts in Problem Representation; Guides to Problem Solving. The session developed a general foundation for problem solving, emphasizing it as a human and humane enterprise. Beginning with the anatomy of a problem, it was shown how culture and human values enter all stages of the problem solving activity. Measurement tools and language were discussed in the context of problem representation, considering the initial state, the goal state, and the search for ways to bridge the gap between the two. The session concluded with a discussion of general percepts in problem representation and guides to the search for solutions.

FOURTH SESSION OF THE ASSOCIATION

The Fourth Session of the Association was a special session organized by Professors Lenore Blum and David Gale entitled "What is Computational Complexity Theory?" The speakers and titles were as follows:

"Algorithms That Toss Coins", Professor Richard Karp, University of California, Berkeley.

"Complexity, Combinatorics, and Group Theory -- A Look at Graph Ismorphism", Dr. Maria M. Klawe, IBM-San Jose.

"Complexity Theory in Number Theory", Professor Leonard Adleman, University of Southern California.

SPECIAL SESSIONS OF THE ASSOCIATION

MAA Mini-Course: "Topics in Data Analysis", by Jon Kettenring and Paul Tukey, Bell Telephone Laboratories.

"Real world" data sets cannot always be readily analysed by the methods of classical statistics, methods which emphasize formal optimality under assumptions which cannot always be verified. When standard techniques do not apply, data analysis can be akin to detective work, hunting for clues to discover buried relationships and structure. Many investigative tools can be applied to data drawn from varied disciplines; some of them are simple enough to be taught to people without mathematical or statistical expertise. Several such methods were taught in this mini-course and were applied to "real world" data.

Film Showings: Film showings were held on Thursday and Saturday evening from 7:00-9:30 P.M.:

Thursday
7:00-8:03 Nim and Other Oriented Graph Games (Gleason)

8:07-8:17	Conics
8:20-8:45	The Gauss-Bonnet Theorem
8:48-9:07	Regular Homotopies in the Plane: Part II
9:11-9:30	Sampling and Estimation, Inferential Statistics, Part I
Saturday	
7:00-7:25	Space Filling Curves
7:28-7:39	Mathematical Peep Show
7:42-7:53	Equidecomposable Polygons
7:56-8:21	Group Theory-A B.B.C. broadcast as part of the Open University's Foundation Course in Mathematics
8:24-8:31	Congruent Triangles
8:34-8:43	Matrices
8:45-8:58	Geometric Introduction to Partial Derivatives
9:02-9:19	Isn't That the Limit
9:22-9:30	Accidental Nuclear War

MEETING OF THE BOARD OF GOVERNORS

The Board of Governors met on Thursday, January 8, 1981, at 9:00 A.M. in the California Room of the San Francisco Hilton with 41 members present. Among the items of business transacted were the following:

The Board elected Professor William G. Chinn, of San Francisco State College, as Second Vice-President of the Association for the two-year term 1981-82. Also elected were Martha Zelinka, Weston High School, and Herbert S. Wilf, University of Pennsylvania, as Governors-at-Large for the three-year term 1981-83.

The Board approved following as members of the Nominating Committee for 1981: Henry L. Alder, Chairman, James L. Cornette, Anne F. O'Neill.

The Board voted to accept the following grants:

1. $9000 for the Hartford, Connecticut region of "Blacks and Mathematics" by Connecticut Mutual Life Foundation ($3000 per year for 3 years).
2. $12,500 for support of "Women and Mathematics" (WAM) by IBM.
3. $5,000 from the Tektronix Foundation for support of the Oregon Region of WAM.
4. A grant of $X from Y for support of the International Mathematical Olympiad of 1981:

X	Y
$ 100	Texas Instruments Coporation
750	Rockwell International Corporation
8,000	International Business Machines Corporation
10,000	Committee on High School Contests

The Board approved the adoption of the following dues schedule for 1982:

	M	M+G	M+T	T	T+G	M+G+T	
Regular	40	50	50	30	40	60	M=AMERIAN MATHEMATICAL MONTHLY
1st 2 yrs.	34	44	44	24	34	54	
Student/Unempl.	20	25	25	15	20	30	
Emeritus	20	25	25	15	20	30	G=MATHEMATICS MAGAZINE
Contributing	80	90	90	70	80	100	
Sponsor	120	130	130	110	120	140	T=TWO-YEAR COLLEGE MATHEMATICS JOURNAL
Patron	200	210	210	190	200	220	
Family			$13 No Journals				
Life	$500 Regular				$1000 Patron		

Also approved was the following schedule of institutional membership dues: Academic Member $125, University Member $225.

ANNUAL BUSINESS MEETING OF THE ASSOCIATION

The Annual Business Meeting of the Association was held on Sunday, January 11, 1981, at the San Francisco Hilton, with Professor Bernstein presiding.

The Association's Nineteenth Award for Distinguished Service was made to Professor Ralph P. Boas of Northwestern University. The citation (which appears on pages 85-86 in the February issue of the MONTHLY) was prepared by Professor Saunders MacLane. Professor Boas, in accepting the Award, spoke as follows:

"First I thank the DSA Committee, the Board of Governors, and the Association at large."

"I did not immediately understand why they asked Saunders Mac Lane to write the citation, but evidently somebody was charmed by the symmetry of the situation -- six years ago Saunders got the Award and I wrote the citation."

"A year or so ago I was at a chamber music concert. As usual, the printed program had biographic notes about the soloist, who in this case was the president of her professional organization, the editor of a newsletter, and so on -- It suddenly struck me that I knew a lot of people like that in the MAA -- they aren't necessarily the most productive mathematicians, but they keep the organization functioning. I think that I am being singled out today as a representative member of that equivalence class, a class that I am happy to think that I belong to."

"Doing mathematics is fun; doing something for mathematics is work; doing it and being appreciated is a real thrill."

A bound copy of the citation authored by Professor Mac Lane was presented to Professor Boas.

The Chauvenet Prize for 1981 was presented to Professor Kenneth I. Gross of the University of North Carolina for his paper, "On the Evolution of Noncommutative Harmonic Analysis," MONTHLY (88), 86-88. Professor Gross, in accepting the Prize, had the following to say:

"Thank you very much. I am grateful to the MAA for such a high honor."

"When I was asked by the MAA in 1977 to lecture on the subject of noncommutative harmonic analysis, and when I wrote this paper that elaborated upon the theme of the talk, for a general audience, I wanted to do a little bit of a lot of things that I had never done before. More than an overview of a subject to which I have devoted the large part of my professional life, I wanted to convey a sense of the history as well as the depth and breadth of an important area of mathematics, along with glimpses of the genius together with the humor that the historian or humanist can find outside the technical field of vision. This was a very enriching experience for me personally, and to have my work so well received has been especially gratifying. To sum up, I feel much like the janitor at the women's college who had a passkey to every room, and who never came by the business office to be paid. When finally asked to pick up his checks, he replied, 'You mean I get paid too?'"

"To turn to a serious vein, I want to thank Ralph Boas, MONTHLY Editor, for his careful reading of the manuscript and a number of thoughtful comments that greatly improved the exposition. I also want to take this opporutnity to reflect upon certain individuals who significantly influenced my career. My teacher, now colleague and friend, Ray A. Kunze, first introduced me to the subject of harmonic analysis. Through his patience and clarity of mathematical expression I was inspired to become a serious research mathematician. From my brother, Herbert I. Gross, a gifted mathematics educator and philosopher, currently at Bunker Hill Community College and M.I.T., I learned the importance of human values, as well as a sense of humor, in the art of teaching. By his model I have tried to elevate the transmission of information, characteristic of our profession, to higher forms of education and human enrichment. In my friend and colleague Ernst Snapper, to whom this Chauvenet paper was dedicated, the aspiring young mathematician could see the rich and varied academic career available for those who apply high standards of excellence in all their professional activities. He is a fountain of eternal mathematical youth. Finally, while in graduate school I had the privilege of a close personal relationship with a professor, Guido Weiss, well-known as an outstanding mathematical expositor. It is a pleasure to follow him, years later, in receiving the Chauvenet Prize. To these people, and many others unnamed, my debt is great." Finally, I have shared my life with three beautiful people, my wife Mary Lou and daughters Laura and Karen, whose love and companionship make this moment especially sweet. Jointly with them, I gratefully accept this award."

ACADEMIC MEMBERS ELECTED BY THE ASSOCIATION

At its meeting on January 8, 1981, the Board of Governors elected as institutional members:

The Kings College, Briarcliff Manor, NY
Middlesex County College, Edison, NJ
York College of CUNY, Jamaica, NY

Respectfully Submitted,
David P. Roselle, Secretary

MATHEMATICAL ASSOCIATION OF AMERICA
Officers and Committees
February 1, 1981

General Offices: Dolciani Mathematical Center
1529 Eighteenth Street, N.W., Washington, D. C. 20036
Telephone 202-387-5200

Executive Director: Alfred B. Willcox
Associate Director: Marcia P. Sward
Editorial Director: Raoul Hailpern
Secretary Emeritus: Henry L. Alder

OFFICERS

President, Richard D. Anderson, Louisiana State University (1981-82)
Past-President, Dorothy L. Bernstein, Applied Mathematics, Brown University (1981-82)
First Vice-President, Lynn A. Steen, St. Olaf College (1980-81)
Second Vice-President, William G. Chinn, San Francisco City College (1981-82)
Editor, Ralph P. Boas, Northwestern University (1977-81)
Secretary, David P. Roselle, Virginia Tech (1980-84)
Treasurer, Leonard Gillman, University of Texas (1978-82)

ADDITIONAL MEMBERS OF THE BOARD OF GOVERNORS

Past-Presidents:

Henry L. Alder, University of California, Davis (1977-85)
Henry O. Pollak, Bell Telephone Laboratories (1975-83)

Elected Members of the Finance Committee:

Donald L. Kreider, Dartmouth College (1978-81)
G. Baley Price, University of Kansas (1980-83)

Editor of MATHEMATICS MAGAZINE: Doris J. Schattschneider, Moravian College (1981-85)

EDITOR OF TWO-YEAR COLLEGE MATHEMATICS JOURNAL: Donald J. Albers, Menlo College (1979-83)

Governors-at-Large:

Peter L. Duren, University of Michigan (1979-81)
Gloria F. Gilmer, Atlanta University (1979-81)
Susan J. Devlin, Bell Telephone Laboratories (1980-82)
Richard K. Guy, University of Calgary (1980-82)
Herbert S. Wilf, University of Pennsylvania (1981-83)
Martha Zelinka, Weston High School (1981-83)

Sectional Governors (July 1, 1978-June 30, 1981):

Allegheny Mountain, Melvin R. Woodard, Indiana University of Pennsylvania
Indiana, Gary J. Sherman, Clemson University
Kentucky, Jacqueline C. Moss, Paducah Community College
Metropolitan New York, Harold N. Shapiro, New York University
Nebraska, Gary H. Meisters, University of Nebraska
Northern California, Kenneth R. Rebman, California State University at Hayward
Oklahoma-Arkansas, Paul E. Long, University of Arkansas
Rocky Mountain, A. Duane Porter, University of Wyoming
Wisconsin, Gary B. Klatt, University of Wisconsin, Whitewater

Sectional Governors (July 1, 1979-June 30, 1982):

Kansas, John J. Hutchinson, Wichita State University
Missouri, Troy L. Hicks, University of Missouri, Rolla
New Jersey, Michael I. Aissen, Rutgers University
Northeastern, Anne F. O'Neill, Wheaton College
Ohio, Samuel W. Hahn, Wittenberg University
Pacific Northwest, Ivan Niven, University of Oregon
Seaway, Mabel D. Montgomery, SUC at Buffalo
Southwestern, Gerald S. Rogers, New Mexico State University
Southeastern, John D. Neff, Georgia Tech

Sectional Governors (July 1, 1980-June 30, 1983):

Florida, Beverly L. Brechner, University of Florida

Illinois, Gordon D. Mock, Western Illinois University
Intermountain, C. Edmund Burgess, University of Utah
Iowa, William L. Waltmann, Wartburg College
Louisiana-Mississippi, David E. Cook, University of Mississippi
Maryland-D.C.-Virginia, Ronald M. Davis, Northern Virginia Community College
Michigan, Delia Koo, Eastern Michigan University
North Central, Dale E. Varberg, Hamline University
Eastern Pennsylvania-Delaware, Gerald J. Porter, University of Pennsylvania
Southern California, John Todd, Caltech
Texas, John T. Mohat, North Texas State University

COMMITTEES OF THE ASSOCIATION

Terms of members expire, except where otherwise noted, at the Annual Meeting in January following the last year of service listed below. For temporary committees, no terms are listed since they are automatically discharged at the expiration of the President's term of office, which is the Annual Meeting in January, 1983.

EXECUTIVE COMMITTEE

Richard D. Anderson, Chairman (1981-82); Dorothy L. Bernstein (1979-81), Ralph P. Boas (1977-81), Leonard Gillman (1978-82), William G. Chinn (1981-82), David P. Roselle (1980-84), Lynn A. Steen (1980-81), all are ex-officio.

FINANCE COMMITTEE

Richard D. Anderson, Chairman (1981-82), ex-officio; Dorothy L. Bernstein (1981-82), ex-officio, Leonard Gillman (1978-82), ex-officio, Donald L. Kreider, (1978-81), David P. Roselle (1980-84), ex-officio, G. Baley Price (1980-83).

Audit and Budget Committee: Donald L. Kreider, Chairman (1978-81); G. Baley Price, (1980-83).

Investment Committee: Leonard Gillman, Chairman (1978-82), ex-officio; Henry L. Alder (1981-83), Edward A. Cameron (1979-81), Harley Flanders, (1981-83), David P. Roselle (1980-84), ex-officio, James M. Vaughn, Jr. (1981-83), Howard E. Zink, (1980-82).

Staff and Services Committee: Leonard Gillman, Chairman (1979-82); Richard D. Anderson, (1981-82), both ex-officio.

ADVISORY COMMITTEE ON THE ARCHIVES OF AMERICAN MATHEMATICS

Judith V. Grabiner, Chairman, (1979-82); Leonard Gillman (1979-82), Robert E. Greenwood (1979-82), Albert C. Lewis, ex-officio, Lucille E. Whyburn (1981-83).

ad hoc COMMITTEE ON CONTINUING EDUCATION

John O. Riedl, Chairman; Robert Bumcrot, Sanford L. Segal.

COMMITTEE ON ADVISEMENT AND PERSONNEL

Bernice L. Auslander, Chairman (1980-81); Jane M. Day (1981-83), Gordon Raisbeck (1980-81), Martha K. Smith (1980-81), William A. Stannard (1981-83), James W. Vick (1980-82).

COMMITTEE ON CORPORATE MEMBERS

Richard A. Moynihan, Chairman (1981-82); Clayton V. Aucoin, (1981-83), David A. Birnbaum (1979-81), Jeremiah J. Lyons (1981-83), Stephen H. Quigley (1980-82), John R. Rasmussen (1979-81), James M. Vaughn (1980-82), Alfred B. Willcox (1980-82).

COMMITTEE ON EARLE RAYMOND HEDRICK LECTURES

F. Burton Jones, Chairman (1981); George E. Andrews (1981-83), Mary Ellen Rudin (1980-82).

COMMITTEE ON HIGH SCHOOL CONTESTS

Stephen B. Maurer, Chairman (1981-83); Ralph A. Artino, (1980-82), George Berzsenyi (1980-82), Richard T. Driver (1981-83), Samuel L. Greitzer (1981-83), Murray S. Klamkin (1981-83), John R. Linden (1980-82), Walter E. Mientka, Director (1979-82), Mary A. Norton (1979-81), Richard S. Pieters (1981-83), Stanley Rabinowitz (1979-81), Alexander Scheitlin (1980-82), Leo J. Schneider, (1980-82), Harold N. Shapiro (1981-83), Nura D. Turner (1980-82), Melvin R. Woodard (1981-83), Martha Zelinka (1980-82).

ADVISORY PANEL OF THE COMMITTEE ON HIGH SCHOOL CONTESTS

Thomas R. Butts, Noel A. Childress, W. C. Foreman, Hubert L. Hunzeker, Frank T. Kocher, Abraham Schwartz, John H. Staib.

SUBCOMMITTEE ON THE USA MATHEMATICAL OLYMPIAD

Samuel L. Greitzer, Chairman (1980-82); George Berzsenyi (1980-82), Murray S. Klamkin (1980-82), Andrew Liu (1981-83), Stephen Maurer (1981-83), ex-officio, Mary A. Norton (1979-81), Cecil C. Rousseau (1979-81).

Advisor to the Olympiad Awards Ceremony: Nura D. Turner (1979-81).

COMMITTEE ON NATIONAL AWARDS AND PUBLIC REPRESENTATION

Peter J. Hilton, Chairman (1980-82); Richard D. Anderson (1981-82), ex-officio, Lester H. Lange (1981-83), Robert L. Wilson (1980-82).

COMMITTEE ON PUBLICATIONS

Edwin F. Beckenbach, Chairman (1980-82); Donald J. Albers (1979-83), ex-officio, Lida K. Barrett (1981-83), Ralph P. Boas (1979-81), ex-officio, Daniel T. Finkbeiner, II (1981-83), Ross L. Finney (1980-82), Leonard Gillman (1978-82), ex-officio, Robert Gilmer (1980-82), Ross A. Honsberger (1980-82), Ivan Niven (1980-83), Barbara L. Osofsky (1981-83), Warren Page (1981-83), Doris W. Schattschneider (1981-85), ex-officio, Alan C. Tucker (1979-81).

Subcommittee on Basic Library Lists: Daniel T. Finkbeiner, II, Chairman; Donald W. Bushaw, H. E. Hall, Robert H. McDowell, Frank L. Wolf. Subcommittee on Carl B. Allendoerfer Awards: Roy Dubisch, Chairman (1981-83); Edwin F. Beckenbach (1980-82), Thomas W. Tucker (1979-81).

Subcommittee on Carus Monographs: Barbara L. Osofsky, Chairman (1981-83); Ralph P. Boas (1979-81), Daniel T. Finkbeiner (1981-83), Robert Gilmer (1980-82).

Subcommittee on Dolciani Mathematical Expositions: Ross A. Honsberger, Chairman (1981-83); Gerald L. Alexanderson (1981-82), Joseph E. Malkevitch (1981-83), Kenneth R. Rebman (1981-82).

Subcommittee on Lester R. Ford Awards: Branko Grunbaum (1981), Chairman; Edwin F. Beckenbach (1980-82), ex-officio.

Subcommittee on MAA Studies in Mathematics: Alan C. Tucker, Chairman (1980-81); Donald W. Anderson (1981-83), Guido L. Weiss (1979-81).

Subcommittee on Miscellaneous Publications: Edwin F. Beckenbach, Chairman (1980-82); Leonard Gillman (1978-82), David P. Roselle (1980-84), all ex-officio.

Subcommittee on the New Mathematical Library: Ivan Niven, Chairman (1981-83); William G. Chinn (1980-82), Basil Gordon (1980-82), Max M. Schiffer (1979-81). Anneli Lax, ex-officio.

Editor of the New Mathematical Library: Anneli Lax.

Subcommittee on George Polya Awards: Kay W. Dundas, Chairman (1980-81); Edwin F. Beckenbach (1980-82), Warren Page (1980-82).

COMMITTEE ON SECONDARY SCHOOL LECTURES

Donald B. Small, Chairman (1979-81); Edward Z. Andalafte (1980-82), Marjorie M. Enneking (1979-81), Gloria F. Gilmer (1980-82), John M. Jobe (1981-83), William K. McNabb (1980-82).

COMMITTEE ON SECTIONS

Samuel W. Hahn, Chairman (1981-84); Yousef Alavi (1980-83), David W. Ballew (1981-84), John D. Neff (1979-82), Kenneth R. Rebman (1981-84), Alfred B. Willcox, ex-officio.

COMMITTEE ON SPECIAL FUNDS OF THE ASSOCIATION

G. Baley Price, Chairman (1981-83); Lida K. Barrett (1981-83), Joshua Barlaz (1980-82), Edward A. Cameron (1979-81), Leonard Gillman (1978-82), ex-officio, Burton W. Jones (1979-81), William A. Stannard (1981-83), Wyman L. Williams (1979-81).

COMMITTEE ON THE AWARD FOR
DISTINGUISHED SERVICE TO MATHEMATICS

Dorothy L. Bernstein, Chairman (1981); Henry L. Alder (1980-82), Ralph P. Boas (1981-83).

COMMITTEE ON THE CHAUVENET PRIZE

Paul R. Halmos, Chairman (1981); Philip J. Davis (1980-82), David Eisenbud (1981-83).

COMMITTEE ON THE EXCHANGE OF INFORMATION ON MATHEMATICS

James R. Leitzel, Chairman (1979-81); Allen L. Hammond, Consultant (1979-81), Yousef Alavi (1981-83),

Ronald M. Davis (1981-83), Donald M. Hill (1981-83), Gary H. Meisters (1981-1983), Jean J. Pedersen (1981-83), Lynn A. Steen (1979-81), Marion D. Wetzel (1980-82).

COMMITTEE ON THE MEMBERSHIP OF THE ASSOCIATION

Dorothy L. Bernstein, Chairman (1981-83); Ronald M. Davis (1980-82), Richard D. Anderson (1981-82), ex-officio, Leonard Gillman (1978-82), ex-officio, Kenneth R. Rebman (1980-82), David P. Roselle, (1980-84), ex-officio, Gary J. Sherman (1979-81), Sue H. Whitesides (1981-83), Paul R. Zuckerman (1980-82).

COMMITTEE ON PLACEMENT EXAMINATIONS

Bernard L. Madison, Chairman (1980-82); Thomas A. Carnevale (1980-82), John W. Kenelly (1980-82), Richard H. Prosl (1981), Billie A. Rice (1981-83), Marcia P. Sward (1979-81).

COMMITTEE ON THE PUTNAM PRIZE COMPETITION

Kenneth B. Stolarsky, Chairman (1981); Gerald L. Alexanderson, Associate Director (1980-82), William Firey (1981-82), Douglas A. Hensley (1981-83), Abraham P. Hillman, Associate Director (1978-82), Leonard F. Klosinski (1980-82).

COMMITTEE ON THE TEACHING OF UNDERGRADUATE MATHEMATICS

James W. Vick, Chairman (1980-81); Henry L. Alder (1979-81), Donald W. Bushaw (1980-82), Gloria F. Gilmer (1981-83), Leon W. Rutland (1979-81), Alan H. Schoenfeld (1979-82), Martha Siegel (1981-82), David I. Schneider (1981-83), June P. Wood (1981-83).

COMMITTEE ON TWO-YEAR COLLEGES

John D. Bradburn, Chairman (1981-83); Donald J. Albers (1980-82), Thomas A. Carnevale (1979-81), Kay W. Dundas (1979-81), Jacqueline C. Moss (1979-81), Warren Page (1980-82), Elaine B. Pavelka (1979-81), Richard H. Plagge (1980-82), Elaine L. Tatham (1981-83), Howard E. Zink (1979-81).

COMMITTEE ON THE UNDERGRADUATE PROGRAM IN MATHEMATICS

Donald W. Bushaw, Chairman (1979-81); Gerald L. Alexanderson (1980-82), Richard D. Anderson (1981-82), ex-officio, Winifred A. Asprey (1980-82), Joseph E. Cicero (1979-81), Jerome A. Goldstein (1980-82), William F. Lucas (1979-81), Bruce E. Meserve (1980-82), Fred S. Roberts (1979-81), Donald B. Small (1980-82), Alan C. Tucker (1980-81).

COMMITTEE ON VISITING LECTURERS AND CONSULTANTS

Malcolm W. Pownall, Chairman (1980-82); Theodore S. Chihara (1981-83), Vivian Y. Kraines (1981-83), Jon M. Laible (1980-82), Eugene R. Seelbach (1979-81).

COMMITTEE ON WORLD WAR II HISTORY

G. Baley Price, Chairman; Churchill Eisenhart, Carroll V. Newsom, Mina S. Rees, J. Barkley Rosser, Raymond L. Wilder. (The terms of all members expire in January, 1984.)

JOINT COMMITTEES

JOINT (MAA-SIAM) COMMITTEE ON APPLICATIONS
OF MATHEMATICS IN THE COLLEGE CURRICULUM

Fred S. Roberts, Chairman (1981); Lily E. Christ (1983, MAA), Joseph E. Flaherty (1982, SIAM)), Peter J. Hilton (1981, MAA), Eugene Isaacson (1982, MAA), John Morrison (1982, SIAM), Thomas G. Proctor (1981, SIAM).

JOINT COMMITTEE ON EMPLOYMENT OPPORTUNITIES

Wilfred E. Barnes, Chairman (1981, MAA); James W. Daniel (1981-83, SIAM), Roger A. Horn (1978-82, AMS), Calvin T. Long (1981-83, MAA), Ervin Y. Rodin (1978-82, SIAM), Donald C. Rung (1981-82, AMS).

JOINT COMMITTEE ON WOMEN IN MATHEMATICS

Mary W. Gray, Chairman (1981), Etta Z. Falconer (1982, AMS), Israel N. Herstein (1981, AMS), Pamela Cook-Ioannides (1983, SIAM), Cathleen S. Morawetz (1981, SIAM), Jacqueline C. Moss (1981, MAA), Edith H. Luchins (1982, MAA), Katherine Pedersen (1983, NCTM), Joel Schneider (1983, NCTM), Barbara Searle (1983, NCTM), Gail A. Williams (1983, MAA).

JOINT MEETINGS COMMITTEE

William J. LeVeque, Chairman; Everett Pitcher, David P. Roselle, Alfred B. Willcox, all ex-officio.

AMS-MAA-SIAM JOINT PROJECTS COMMITTEE FOR MATHEMATICS
(Terms expire September 30)

Felix E. Browder, Chairman (1981); C. Edmund Burgess (1983), Wendell H. Fleming (1982), Shirley A. Hill (1983), Donald L. Kreider (1981), Gottfried E. Noether, (1982), Seymour V. Parter (1982), Werner C. Rheinboldt (1981).

CEEB-MAA COMMITTEE ON MUTUAL CONCERNS

Henry L. Alder (1982, MAA), Richard D. Anderson (1983, MAA), Lida K. Barrett (CEEB), Chancey O. Jones (CEEB), Donald L. Kreider (CEEB), David P. Roselle, Co-Chairman (1981, MAA), Howard E. Taylor, Co-Chairman (CEEB), Alfred B. Willcox, ex-officio (MAA).

EDITORIAL BOARDS OF THE ASSOCIATION

AMERICAN MATHEMATICAL MONTHLY
(All terms expire December 31, 1981)

Editor: Ralph P. Boas, Northwestern University
Editor-Elect: Paul R. Halmos, Indiana University

Associate Editors: Ezra A. Brown, Vladimir Drobot, Martha W. Evens, David Gale, Richard K. Guy, Raoul Hailpern, Deborah T. Haimo, Franklin T. Haimo, Frank T. Kocher, Jr., Abraham P. Hillman, R. C. Lyndon, M. S. Montgomery, Timothy J. Robertson, Seymour Schuster, J. A. Seebach, Jr., Ivar Stakgold, Emory P. Starke, Lynn A. Steen, Alan C. Tucker, Mary R. Wardrop, Robert F. Wardrop.

MATHEMATICS MAGAZINE
(All terms expire December 31, 1985)

Editor: Doris J. Schattschneider, Moravian College

Associate Editors: Edward J. Barbeau, John A. Beidler, Paul J. Campbell, Underwood Dudley, Dan J. Eustice, Joseph A. Gallian, Judith V. Grabiner, Raoul Hailpern, Pierre J. Malraison, Joseph Malkevitch, Leroy F. Myers, Jean J. Pedersen, Gordon Raisbeck, Ian Richards, David A. Smith.

TWO-YEAR COLLEGE MATHEMATICS JOURNAL
(All terms expire December 31, 1983)

Editor: Donald J. Albers, Menlo College

Associate Editors: Gerald L. Alexanderson, Glenn D. Allinger, Anthony Barcellos, William G. Chinn, Ronald M. Davis, Howard M. Eves, Stanley Friedlander, Thomas M. Green, Samuel A. Greesnpan, Raoul Hailpern, Ross A. Honsberger, Harold R. Jacobs, Erwin Just, Bruce W. King, Norman E. Ladd, Roland H. Lamberson, William W. Leonard, Peter A. Lindstrom, David A. Logothetti, Nancy Myers (NCTM), Warren Page, George Polya, Edward B. Wright.

NEWSLETTER EDITORIAL COMMITTEE
(All terms expire December 31, 1985)

J. Arthur Seebach, Jr., Co-Chairman; Lynn A. Steen, Co-Chairman; Henry L. Alder, Philip M. Cheifetz, Ronald M. Davis, Katherine P. Layton, James R. Leitzel, ex-officio, Eileen L. Poiani, Kenneth R. Rebman, David P. Roselle, ex-officio, Marcia P. Sward, ex-officio, Alan C. Tucker.

REPRESENTATIVES OF THE ASSOCIATION

On Sections of AAAS:
Section A: Lowell J. Paige (1981-83)
Section Q: Joan P. Leitzel (1981-83)
Section U: Fred Mosteller (1981-83)
Section X: Alfred B. Willcox (1981-83)

On UMAP Editorial Board: Robert G. Bartle (1981-82), Anne F. O'Neill (1981-82), Peter A. Lindstrom (1981-82)

On the Council of the Conference Board of the Mathematical Sciences:
Dorothy L. Bernstein, ex-officio, Donald L. Kreider, ex-officio

On the Discipline Committee for Mathematics of CEEB: Donald L. Kreider

On the Governing Council of Mu Alpha Theta: Robert L. Wilson (1979-81)

On the U. S. Commission on Mathematical Instruction:
Henry L. Alder (July 1, 1977-June 30, 1981), Leon A. Henkin (July 1, 1978-June 30, 1982)

OFFICERS OF THE SECTIONS

ALLEGHENY MOUNTAIN

Chairman: Allan Krall, Pennsylvania State University, University Park
1st Vice-Chairman: Francis H. S. Hall, Pennsylvania State University, Fayette Campus
2nd Vice-Chairman: Donald M. Platte, Mercyhurst College
Coordinator of Student Programs: Barbara Faires, Westminster College
H.S. Contest Regional Coordinator: Nicholas Ford, Pennsylvania State University, Fayette (W. PA)
I. Dee Peters, University of West Virginia (W. VA)
Newsletter Editor: Kathleen Taylor, Duquesne University
Secretary/Treasurer: John M. Atkins, West Virginia University, Morgantown

EASTERN PENNSYLVANIA AND DELAWARE

Chairman: Howard Anton, Drexel University
Vice-Chairman: Bing Wong, Wilkes College
H.S. Contest Regional Coordinator: Albert E. Filano, West Chester State College (E. PA)
Jack Fink, Continental American Life Insurance Co. (DE)
Newsletter Editor: Paul Cochrane, Bloomsburg State College
Secretary/Treasurer: Willard E. Baxter, University of Delaware

FLORIDA

Chairman: Edwin Duda, University of Miami
Chairman-Elect: Robert Gilmer, Florida State University, Tallahassee
Vice-Chairman: Ernest Ross, St. Petersburg Jr. College, Clearwater
Vice-Chairman: Raymond E. Roth, Rollins College
Past-Chairman: Frederick Hoffman, Florida Atlantic University
H.S. Contest Regional Coordinator: William Hoop, American Bankers Assurance Co. (FL)
Walter E. Mientka, University of Nebraska, Lincoln (Canal Zone)
Eugene A. Francis, University of Puerto Rico (PR)
Newsletter Editor: Edwin Duda, University of Miami
Secretary/Treasurer: Frank L. Cleaver, University of South Florida, Tampa

ILLINOIS

Chairman: Robert N. Pendergrass, Southern Illinois University
Chairman-Elect: John S. Haverhals, Bradley University
1st Vice-Chairman: Ronald M. Shelton, Milliken University
2nd Vice-Chairman: Barbara Juister, Elgin Community College
H.S. Contest Regional Coordinator: Gary Tippett, Bradley University
Newsletter Editor: John Hooker, Southern Illinois University
Secretary/Treasurer: Howard C. Saar, 216 Willow Drive East, Plainfield, IL

INDIANA

Chairman: Duane E. Deal, Ball State University
Vice-Chairman: Marvin G. Mundt, Valparaiso University
H.S. Contest Regional Coordinator: Stephen Terry, The Nyhart Co.
Newsletter Editor: Douglas R. McCarthy, Indiana-Purdue University, Fort Wayne
Secretary/Treasurer: Richard R. Patterson, Indiana-Purdue University, Indianapolis

INTERMOUNTAIN

Chairman: Richard Tebbs, Southern Utah State College
1st Vice-Chairman: Richard R. Miller, Weber State College
2nd Vice-Chairman: Linda C. Hill, Idaho State University
H.S. Contest Regional Coordinator: Robert Vunovich, University of S. Colorado (UT)
Tom Williams, Idaho State University (ID)
Secretary/Treasurer: Lawrence O. Cannon, Utah State University, Logan

IOWA

Chairman: Arnold M. Adelberg, Grinnell College
Vice-Chairman: A. M. Fink, Iowa State University, Ames
H.S. Contest Regional Coordinator: Charles Smith, The Bankers Life
Secretary/Treasurer: James Peake, Iowa State University, Ames

KANSAS

Chairman: Thomas S. Parker, Kansas State University, Manhattan

Vice-Chairman: Sr. Jo Ann Fellin, Benedictine College
Associate Chairman for Community Colleges: Kay W. Dundas, Hutchinson Community College
H.S. Contest Regional Coordinator: Steven L. Cooper, Security Benefit Life
Secretary/Treasurer: Robert H. Thompson, Washburn University

KENTUCKY

Chairman: Kyle Wallace, Western Kentucky University
Chairman-Elect: Richard Davitt, University of Louisville
Vice-Chairman: Edgar E. Enochs, University of Kentucky, Lexington
H.S. Contest Regional Coordinator: William Parker, Jefferson Community College, S.W.
Newsletter Editor: Carroll Wells, Western Kentucky University
Secretary/Treasurer: Joe K. Smith, Northern Kentucky University

LOUISIANA-MISSISSIPPI

Chairman: Edwin P. Oxford, University of Southern Mississippi
LA Vice-Chairman: James H. Abbott, University of New Orleans
MS Vice-Chairman: Daniel A. Hogan, Hinds Junior College
H.S. Contest Regional Coordinator: Benjamin Mitchell, Louisiana State University (LA)
David E. Cook, University of Mississippi (MS)
Newsletter Editor: John L. Tilley, Mississippi State University
Secretary/Treasurer: John L. Tilley, Mississippi State University

MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA

Chairman: John M. Smith, George Mason University
Chairman-Elect: John Schmeelk, Virginia Commonwealth University
Vice-Chairman of Programs: Patrick Hayes, 950 25th St., N.W., Apt. 707N, Washington, D.C. 20037
Vice-Chairman of Membership: Howard L. Penn, U.S. Naval Academy
Treasurer: Arthur T. Charlesworth, University of Richmond
H.S. Contest Regional Coordinator: Edward D. Bender, J. Sargeant Reynolds Community College
Newsletter Editor: Howard L. Penn, U.S. Naval Academy
Secretary: John R. Hanson, James Madison University

METROPOLITAN NEW YORK

Chairman: Godfrey Isaacs, Herbert H. Lehman College
Vice-Chairman of 4-year Colleges: Theresa J. Barz, St. John's University
Vice-Chairman of 2-year Colleges: Leo M. Levine, Queensborough Community College
Vice-Chairman for High School: Harry Ruderman, Hunter College H.S., 94th & Park Avenue, NY, NY 10028
Treasurer: Robert G. Payton: Adelphi University
H.S. Contest Regional Coordinator: Sandra Pulver, Pace University
Newsletter Editor: William F. Orr, Hofstra University
Secretary/Treasurer: Lily E. Christ, John Jay College of Criminal Justice

MICHIGAN

Chairman: Harold T. Slaby, Wayne State University
Vice-Chairman: David James, University of Michigan, Dearborn
Vice-Chairman: Philip H. Mahler, Henry Ford Community College
H.S. Contest Regional Coordinator: Joanne Rankin, Eastern Michigan University
Newsletter Editor: Delano Wegener, Central Michigan University
Secretary/Treasurer: Paul K. Garlick, University of Michigan, Flint

MISSOURI

President: Merry McDonald, Northwest Missouri State University
Vice-President: A. Glen Haddock, University of Missouri, Rolla
Chairman of H.S. Lecturers: Leonard Palmer, Southeast Missouri State University
Past President: Michael Z. Williams, Westminster College
H.S. Contest Regional Coordinator: Alvin Tinsley, Central Missouri State University
Newsletter Editor: Elizabeth Berman, University of Missouri, Kansas City
Secretary/Treasurer: Jerry Wilkerson, Missouri Western State College

NEBRASKA

Chairman: Alexander Mehaffey, Jr., University of South Dakota
Chairman-Elect: Randall K. Heckman, Kearney State College
H.S. Contest Regional Coordinator: Stanley D. Luke, Nebraska Wesleyan University

Newsletter Editor: H. M. Cox & Thomas S. Shores, University of Nebraska, Lincoln
Secretary/Treasurer: Henry M. Cox, 1145 North 44th Street, Lincoln, NE 68503

NEW JERSEY

Chairman: Jean Lane, Union College
Chairman-Elect: Susan G. Marchand, Kean College
Treasurer: Judith B. Seery, Bell Labs, Murray Hill, NJ 07974
Vice-Chairman of Innovations: Dorothy Wolff, Caldwell College
Vice-Chairman of Speakers: Sr. Stephanie Sloyan, Georgian Court College
Vice-Chairman of TYC: Margaret Piedem, Somerset Community College
Vice-Chairman for H.S. Contest: Francis Masat, Glassboro State College
H.S. Contest Regional Coordinator: Robert Mitchell, Glassboro State College
Newsletter Editor: Jean Lane, Union College
Secretary: James Magliano, Union College

NORTH CENTRAL

President: Stephen R. Hilding, Gustavus Adolphus College
President-Elect: Roger B. Kirchner, Carleton College
Past President: Hubert R. Walczak, College of St. Thomas
Member At Large: Roger A. Avelsgaard, Bemidji State University
Member At Large: Harlan A. Hewitt, North Hennepin Community College
H.S. Contest Regional Coordinator: Wayne Roberts, Macalester College (MN)
John Whitcomb, University of North Dakota (ND)
John Barsby, St. John's Ravenscourt (Manitoba)
Marty Streiker, Evan Hardy Collegiate Institute (Saskatchewan)
Newsletter Editor: Joseph Gallian, University of Minnesota, Duluth
Secretary/Treasurer: Charles V. Heuer, Concordia College

NORTHEASTERN

Chairman: Roger L. Cooke, University of Vermont, Burlington
Vice-Chairman: James E. Ward, Bowdoin College
2-Year Representative: Nancy Myers, 67 Highland Avenue, Cambridge, MA 02139
H.S. Contest Regional Coordinator: Paul Norris, Connecticut Mutual Life Insurance Co. (CT)
Olin Sawyer, Union Mutual Life Insurance Co. (ME)
Murray Abramson, Bridgewater State College (MA)
Owen R. Koppang, Concord High School (NH and RI)
J. S. H. Johansson, University of Vermont (VT)
William Crawford, Mt. Allison University (New Brunswick, Newfdld., Nova Scotia, P.E. Island)
Newsletter Editor: Dorothy T. Meserve, Box 108, Fairfax, VT 05454
Secretary/Treasurer: Shirley A. Blackett, 97 Eliot Avenue, West Newton, MA 02165

NORTHERN CALIFORNIA

Chairman: Sr. Madeleine R. Ashton, College of the Holy Names
Vice-Chairman: Hans Samelson, Stanford University
Program Chairman: Carroll Wilde, Naval Postgraduate School
H.S. Contest Regional Coordinator: Robert McFarland, Alameda Cty Office of Education (No. CA and NV)
Jack Johnson, Brigham Young University (Hawaii, Pacific Islands, Guam)
Newsletter Editor: Henry Osner, Modesto Junior College
Secretary/Treasurer: Leonard F. Klosinski: University of Santa Clara

OHIO

Chairman: Clifford A. Long, Bowling Green State University
Program Chairman: Darrell J. Horwath, John Carroll University
2-Year College Representative: Cynthia P. Yang, Miami University
Past Chairman: Donald O. Koehler, Miami University
H.S. Contest Regional Coordinator: Leo Schneider, John Carroll University (OH)
I. Dee Peters, University of W. Virginia (WV)
Newsletter Editor: Richard A. Little, Baldwin-Wallace College
Secretary/Treasurer: Gus Mavrigian, Youngstown State University

OKLAHOMA-ARKANSAS

Chairman: Andrew D. Coe, Westark Community College

Past Chairman: Jim Choike, Oklahoma State University
1st Vice-Chairman: Robert D. McMillan, Oklahoma Christian College
2nd Vice-Chairman: Allan C. Cochran, University of Arkansas, Fayetteville
H.S. Contest Regional Coordinator: Thomas Cairns, University of Tulsa (OK)
Claude Duplissey, University of Arkansas, Little Rock (AR)
Newsletter Editor: Cecil McDermott and William Murray, Hendrix College
Secretary: John Jobe, Oklahoma State University

PACIFIC NORTHWEST

Chairman: Norman F. Lindquist, Western Washington University
Chairman-Elect: Beryl Green, Oregon College of Education
Vice-Chairman 4-Year Colleges: Marjorie M. Enneking, Portland State University
Vice-Chairman 2-Year Colleges: Louise Dyson, Clark College
H.S. Contest Regional Coordinator: Donval R. Simpson, Institute of Arctic Biology (AK)
Tom Williams, Idaho State University (ID)
William Myers, University of Montana (MT)
Michael N. Dyer, University of Oregon (OR)
Calvin Long, Washington State University (WA)
Harold L. Lampkin, University of Calgary (Alberta)
Leon Bowden, University of Victoria (Brit. Columbia, Yukon, Northwest Territories)
Newsletter Editor: Sue Kaplan, Western Washington University
Secretary/Treasurer: Samuel L. Dunn, Seattle Pacific University

ROCKY MOUNTAIN

Chairman: William C. Ramaley, Ft. Lewis College
Chairman-Elect: John P. Gill, University of Southern Colorado
Program Chairman: Stephen Jay Schiffman, Colorado College
Vice-Chairman: Aubrey Owen, Community College of Denver, No. Campus
H.S. Contest Regional Coordinator: Robert Vunovich, University of Southern Colorado (CO)
William Myers, University of Montana (MT)
Newsletter Editor: David Ballew, South Dakota School of Mines and Technology
Secretary Treasurer: David Ballew, South Dakota School of Mines and Technology

SEAWAY

Chairperson: Howard E. Bell, Brock University, St. Catherines
1st Vice-Chairperson: Jack E. Graver, Syracuse University
2nd Vice-Chairperson: Harriette J. Stephens, S.U.N.Y., College at Canton
Chairperson-Elect: Kenneth D. Magill, S.U.N.Y., Center at Buffalo
H.S. Contest Regional Coordinator: Dennis Martin, S.U.N.Y., College at Brockport (Upper NY)
Louis F. Scholl, Trocaire College (Western NY)
F. A. Sherk, University of Toronto (Ontario)
Saverio Mirarchi, Laval Catholic High School (Quebec)
Newsletter Editor: Donald Trasher, S.U.N.Y., College at Geneseo
Secretary/Treasurer: Donald Trasher, S.U.N.Y., College at Geneseo

SOUTHERN CALIFORNIA

Chairman: Adil Yaqub, University of California, Santa Barbara
1st Vice-Chairman: Judith Grabiner, California State University, Dominguez Hills
2nd Vice-Chairman: Wayne Gibson, Santa Ana College
Program Chairman: Alfred W. Hales, University of California, Los Angeles
Past Chairman: John McGhee, California State University, Northridge
H.S. Contest Regional Coordinator: Dwight O. Coblentz, Department of Education
Newsletter Editor: J. Myron Hood, California Polytech State University
Secretary/Treasurer: J. Myron Hood, California Polytech State University

SOUTHEASTERN

Chairman: Lida K. Barrett, University of Tennessee
Chairman-Elect: James R. Garrett, Lenoir-Rhyne College
Vice-Chairman: Catherine C. Aust, Clayton Jr. College
Lecturer: Daniel D. Warner, Clemson University
H.S. Contest Regional Coordinator: James O'Neal, Troy State University (AL)
Jan List Boal, Georgia State University (GA)
Hal B. Phillips, Jr., Jefferson Standard Life Insurance Co. (NC)
Stanley Lukawecki, Clemson University (SC)
Sherwood Z. Smith, National Life & Accident Insurance Co. (TN)

Newsletter Editor: DeWayne Nymann, University of Tennessee, Chattanooga
Secretary/Treasurer: Ivey C. Gentry, Wake Forest University

SOUTHWESTERN

Chairman: Donald G. Johnson, New Mexico State University
Vice-Chairman: William Y. Velez, University of Arizona
H.S. Contest Regional Coordinator: D. R. Arterburn, New Mexico Institute of Mining & Tech (AZ)
Secretary/Treasurer: Alvin Swimmer, Arizona State University

TEXAS

Chairman: Bill D. Anderson, East Texas State University
Chairman-Elect: Robert L. Tennison, University of Texas, Arlington
Level I Director: David Sanchez, San Antonio College
Level II Director: R. Vic Morgan, Sul Ross State University
Level III Director: Bennie B. Williams, University of Texas, Arlington
Director at Large: Landon Colquitt, Texas Christian University
Arrangements Chair: Raymond W. Tebbetts, San Antonio College
Arrangements Chair-Elect: Richard A. Alo, Lamar University
H.S. Contest Regional Coordinator: James R. Boone, Texas A & M University
Newsletter Editor: Glen E. Mattingly, Sam Houston State University
Secretary/Treasurer: Glen E. Mattingly, Sam Houston State University

WISCONSIN

Chairman: Anthony E. Barkauskas, University of Wisconsin, La Crosse
Chairman-Elect: Karl Beres, Ripon College
H.S. Contest Regional Coordinator: Arthur Brownell, Jr., University of Wisconsin, La Crosse
Newsletter Editor: Robert W. Langer, University of Wisconsin, Eau Claire
Secretary/Treasurer: Paul J. Campbell, Beloit College

CALENDAR OF FUTURE MEETINGS

Sixty-first Summer Meeting, University of Pittsburgh, Pittsburgh, Pennsylvania, August 17–19, 1981.
Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Duquesne University, Pittsburgh, Pennsylvania, May 15–16, 1981.
EASTERN PENNSYLVANIA AND DELAWARE, Villanova University, Villanova, Pennsylvania, November 21, 1981.
FLORIDA, early March. Deadline for paper titles two weeks before meeting.
ILLINOIS, Illinois State University, Normal, May 1–2, 1981.
INDIANA
INTERMOUNTAIN
IOWA, third weekend in April. Deadline for papers February 1.
KANSAS, March or April. Deadline for papers January 1.
KENTUCKY, early April. Deadline for papers six weeks before meeting.
LOUISIANA–MISSISSIPPI, Friday-Saturday before February 20. Deadline for papers three months before meeting.
MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, Saturday before Thanksgiving and last Saturday in April.
METROPOLITAN NEW YORK, Lehman College, CUNY, May 2, 1981.
MICHIGAN, Oakland University, Rochester, May 1–2, 1981.
MISSOURI, late March/early April. Deadline for papers January 31.
NEBRASKA, April.
NEW JERSEY, early November and early May.
NORTH CENTRAL, Mankato State University, Mankato, Minnesota, May 1–2, 1981.
NORTHEASTERN, New England College, Henniker, New Hampshire, June 12–13, 1981.
NORTHERN CALIFORNIA, first or second Saturday in February.
OHIO
OKLAHOMA–ARKANSAS, (approx.) Friday and Saturday of first weekend in April. Deadline for papers three weeks before meeting.
PACIFIC NORTHWEST, Lewis and Clark College, Portland, Oregon, June 19–20, 1981.
ROCKY MOUNTAIN, Colorado College, Colorado Springs, May 1–2, 1981.
SEAWAY, first Saturday in November and Saturday in late April. Deadline for papers six weeks before meeting.
SOUTHEASTERN
SOUTHERN CALIFORNIA, first or second Saturday in March.
SOUTHWESTERN, usually in April. Deadline for papers two weeks before meeting.
TEXAS, Friday and Saturday in early April. Deadline for papers March 1.
WISCONSIN, Friday and Saturday between mid-April and first week in May. Deadline for papers six weeks before meeting.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES
AMERICAN MATHEMATICAL SOCIETY, University of Pittsburgh, Pittsburgh, Pennsylvania, August 18–21, 1981.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.
ASSOCIATION FOR SYMBOLIC LOGIC
ASSOCIATION FOR WOMEN IN MATHEMATICS, University of Pittsburgh, Pittsburgh, Pennsylvania, August 17–21, 1981.
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, April 14–17, 1982.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Four Seasons Sheraton, Toronto, Canada, May 4–6, 1981.
PI MU EPSILON
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.

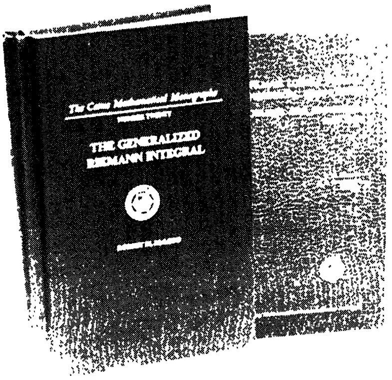
FOUR NEW BOOKS FROM THE MAA...

MAA Studies in Mathematics #20

STUDIES IN ALGEBRAIC GEOMETRY
A. Seidenberg, editor
143 + ix pages hardbound.
List: $16.00 MAA Member: $12.00

STUDIES IN ALGEBRAIC GEOMETRY, a new volume in the MAA Studies in Mathematics Series, contains expository articles which make some of the most interesting results in algebraic geometry accessible to students, as well as teachers and mathematicians who are nonspecialists or scientists working in other fields.

The contents include chapters on "initial results in the theory of linear algebraic groups," "the connectedness theorem and the concept of multiplicity," "space curves as ideal-theoretic complete intersections," and "Chasle's enumerative theory of conics."

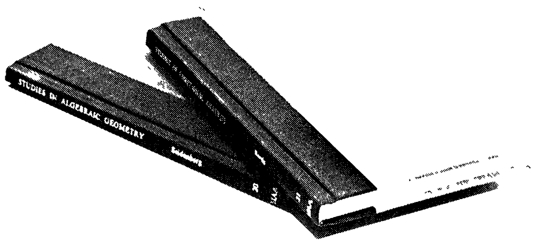


MAA Studies in Mathematics #21

STUDIES IN FUNCTIONAL ANALYSIS
Robert Bartle, editor
227 + xi pages hardbound.
List: $19.00 MAA Member: $14.00

STUDIES IN FUNCTIONAL ANALYSIS, a new addition to the MAA's series STUDIES IN MATHEMATICS, presents five expository articles on recent developments in functional analysis. Although these articles cannot provide a full account of the current state of the field, they do give a reader a tantilizing taste of some of the latest ideas and applications of this fascinating area of mathematics.

The contents include chapters on "numerical ranges," "projection operators in approximation theory," "complementably universal separable banach spaces; an application of counter examples to the approximation problem," "integral representation for elements of convex sets," and "aspects of banach lattices."



Carus Mathematical Monographs #20

THE GENERALIZED RIEMANN INTEGRAL
by Robert M. McLeod
275 + xiii pages hardbound.
List: $18.00 MAA Member: $13.50

THE GENERALIZED RIEMANN INTEGRAL is the first book-length presentation of an exciting new development in integration theory. A new integral, a generalization of the familiar Riemann integral, has been discovered which has all the power and range of the Lebesque integral but which can be readily understood by anyone who has studied calculus through the sophomore level.

Because of the clarity and organization of the exposition and the inclusion of exercises designed to actively engage the reader in the material, it is eminently suitable for use as a textbook at the advanced undergraduate or beginning graduate level.

A New Problem Book from the MAA...

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION: PROBLEMS AND SOLUTIONS: 1938-1964
A. M. Gleason, R. E. Greenwood, and L. M. Kelly
652 + xi pages hardbound.
List: $35.00 MAA Member: $26.00

Problemists the world over, including all past and future Putnam Competitors, will revel in "mastering the difficulties" posed by this collection of problems from the first 25 William Lowell Putnam Competitions. Solutions to all 347 problems are given. In some cases multiple solutions are included, some which contestants could reasonably be expected to find under examination conditions and others which are more elegant or utilize more sophisticated techniques. Valuable references and comments on many of the problems are presented.

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

JUNE-JULY

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 6

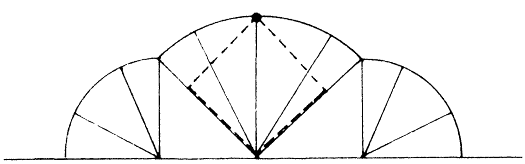
Cutting Lemniscates

Rigid Motions of Conics

Space Filling Curves

Rolling Curves

Iterated Maps



Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88

NUMBER 6

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

ABEL'S THEOREM ON THE LEMNISCATE

MICHAEL ROSEN
Department of Mathematics, Brown University, Providence, RI 02906

Introduction. As is well known, one of the great accomplishments of Gauss's youth was the proof that a regular polygon of 17 sides could be constructed with ruler and compass. In the seventh chapter of *Disquisitiones Arithmeticae* [2] he shows how this result follows from his arithmetic theory of cyclotomic numbers. In general he shows that a regular n-gon can be constructed with ruler and compass if $n = 2^a p_1 p_2 \cdots p_t$ where the p_t are distinct Fermat primes. He also asserts that he has a proof of the converse and warns the reader not to attempt to construct other polygons "and so spend his time uselessly."

In the introduction to Chapter 7 Gauss states that the principles underlying his theory apply not only to circular functions (sine, cosine, etc.) but also to a much larger class of transcendental functions "for example those that depend on the integral $\int dt/\sqrt{1-t^4}$...." He says that he is preparing a comprehensive work on this subject. Unfortunately this treatise never materialized. The functions in question are nowadays called elliptic functions. Gauss's unpublished papers reveal that by 1801, when *Disquisitiones Arithmeticae* appeared, he was already in possession of large parts of elliptic function theory (see [3]). The epoch-making works of Abel and Jacobi were to appear some twenty-five years later.

Gauss kept a mathematical diary [3] from 1796 to 1814. It was not found until 1898, forty-three years after his death. There are 146 entries, all of them short notices of new results. The entry of March 21, 1797, reads as follows, "Lemniscata geometrice in quinque partes dividitur." In other words, he had discovered how to divide the lemniscate into five equal parts with ruler and compass. Among other things this result is remarkable because it shows that at this early date Gauss already knew something about complex multiplication of elliptic functions.

Abel, of course, knew nothing of Gauss's diary. He was, however, very familiar with *Disquisitiones Arithmeticae* and was especially intrigued (as was Jacobi) with the remarks, mentioned above, about "a much larger class of transcendental functions." In 1826, while working on the division equation for elliptic functions, he gained insight into Gauss's theory. He wrote to Holmboe: "On that same occasion I have lifted the mystery which had rested over Gauss' theory of the division of the circle; I now see as clear as daylight how he has been led to it."* His work on elliptic functions was coming along at a rapid pace and he wrote to Crelle and Holmboe with obvious excitement about his forthcoming treatise: "...in which there are many queer things which I flatter myself will startle someone; among others it is about the division of arcs of the lemniscate. You will see how pretty it is!" (For this and the previous quote see Chapter 13 of [6].)

The finished work, "Recherches sur les fonctions elliptiques," appeared in two parts in volumes 2 and 3 of *Crelle's Journal der Mathematik*. The two articles take up 197 pages. They lay the foundations of the theory of elliptic functions and contain a cornucopia of beautiful results. Among these is the following gem, which seems to have been virtually forgotten.

THEOREM. *The lemniscate can be divided into n equal parts with ruler and compass if* $n = 2^a p_1 p_2 \cdots p_t$ *where the* p_i *are distinct Fermat primes.*

---

*I would like to thank Professor H. Edwards of the Courant Institute for bringing this quote to my attention.

Michael Rosen received his Ph.D. from Princeton University in 1963. He has spent most of his academic career at Brown University, where he is now Professor of Mathematics. He has been on leave three times: at Brandeis University, 1965–66; at the University of Wisconsin (Madison), 1971–72; and at the University of California (Berkeley), 1979–80. His main research interests are algebraic number fields and algebraic function fields. He is coauthor with Ken Ireland of *Elementary Number Theory—Including an Introduction to Equations over Finite Fields.*
—Editors

This result is the exact analogue of Gauss's result for the circle. Clearly, it goes far beyond what Gauss recorded in the diary entry of March 21, 1797.

The main object of this paper is to give a reasonably elementary proof of Abel's theorem and its converse. For the most part we use only the beginnings of elliptic function theory (see, for example, the first chapter of [**5**], [**7**], or [**8**]). Our proof differs from that of Abel in that we use Galois theory (which was unavailable to him) and relies on the properties of the Weierstrass $\wp$ function rather than the lemniscate function $\phi(z)$ (to be defined below). Nevertheless the main point of both proofs is the same; the functions involved admit complex multiplication by the ring of Gaussian integers $Z[i]$.

In Section 1 we review the Galois-theoretic proof of Gauss's theorem and recast it in such a way that the later work on lemniscate appears as a natural generalization. In Section 2 we define the lemniscate, discuss its arc-length, define the lemniscate function $\phi(z)$, and discuss the remarkable properties of this function discovered by Abel (and independently by Gauss and Jacobi). In Section 3 we briefly review the elements of the theory of elliptic functions. In Section 4 we relate the function $\phi(z)$ to the Weierstrass $\wp$-function. Finally, in Section 4 we give our proof of Abel's theorem and its converse. To the best of our knowledge, a proof of the converse of Abel's theorem has not previously appeared in print.

In addition to presenting some material of great historical interest we hope this paper will serve as an introduction to the arithmetic theory of elliptic curves, an area of mathematics which is alive and well and being pursued with great intensity by number-theorists of the present day.

1. As is well known, a complex number α is constructible with ruler and compass if and only if $\mathbf{Q}(\alpha)$ is contained in a field K obtained from the rational numbers $\mathbf{Q}$ by a succession of quadratic extensions. It is equivalent to require that α be in a field K which is Galois over $\mathbf{Q}$ and such that $G(K/\mathbf{Q})$, the Galois group of K over $\mathbf{Q}$, has order a power of two.

Let $\zeta_n = \exp(2\pi i/n)$. One knows that $\mathbf{Q}(\zeta_n)/\mathbf{Q}$ is a Galois extension with Galois group isomorphic to $(Z/nZ)^*$. Elementary number theory shows that the order of $(Z/nZ)^*$ is a power of two if and only if $n = 2^a p_1 p_2 \cdots p_t$ where the p_i are distinct Fermat primes. This proves Gauss's theorem.

A slightly different approach goes as follows. Map the real numbers $\mathbf{R}$ to the unit circle C by $\xi(t) = (\cos t, \sin t)$. The map ξ is onto and periodic. The periods consist of all multiples of $2\pi, \langle 2\pi \rangle$. Thus ξ gives rise to a bijection between $\mathbf{R}/\langle 2\pi \rangle$ and C. Since $\mathbf{R}/\langle 2\pi \rangle$ is a group, C can be made into a group by "transport of structure." Using the addition formulae for sine and cosine we see that the group law on C is given by $(a,b) + (c,d) = (ac - bd, ad + bc)$. The unit element of C is $(1,0)$. For n a positive integer we deduce that there are polynomials $f_n(x,y), g_n(x,y) \in Z[x,y]$ such that $n(x,y) = (x,y) + (x,y) + \cdots + (x,y) = (f_n(x,y), g_n(x,y))$. For example, $2(x,y) = (x^2 - y^2, 2xy)$ and $3(x,y) = (x^3 - 3xy^2, 3x^2y - y^3)$. Let C_n be the points of order dividing n on C. Since $C \approx \mathbf{R}/\langle 2\pi \rangle$ we see $C_n \approx Z/nZ$. Moreover,

$$C_n = \{(a,b) \in C \mid f_n(a,b) = 1 \text{ and } g_n(a,b) = 0\}.$$

Let σ be an automorphism of $\mathbf{C}$, the complex numbers, over $\mathbf{Q}$. Since $f_n(a,b)^\sigma = f_n(a^\sigma, b^\sigma)$ and $g_n(a,b)^\sigma = g_n(a^\sigma, b^\sigma)$, we see σ maps C_n into itself. Since C_n is finite, it follows that the coordinates of the points in C_n are algebraic numbers. Adjoin these coordinates to $\mathbf{Q}$ and call the resulting field K_n. From what has been said it follows that $K_n/\mathbf{Q}$ is a Galois extension. Denote the Galois group by G_n. G_n acts on C_n and, since the group law is defined by polynomials with coefficients in $\mathbf{Q}$, we see that G_n preserves the group structure of C_n. Thus we have a map from G_n to $\mathrm{Aut}(C_n)$ which is easily seen to be a homomorphism. The map is actually a monomorphism since the coordinates of the points in C_n generate K_n. Thus G_n is isomorphic to a subgroup of $\mathrm{Aut}(C_n) \approx \mathrm{Aut}(Z/nZ) \approx (Z/nZ)^*$. The proof of Gauss's theorem now follows as before.

This second proof may seem overly elaborate, but, as we shall see, many of the ideas involved will be useful later.

2. A lemniscate may be defined geometrically as the locus of all points such that the product of the distances to two fixed points is a constant. This definition gives rise to a family of curves. We normalize matters by requiring the fixed points to be $(-\sqrt{2}/2,0)$ and $(\sqrt{2}/2,0)$ and the constant to be $\frac{1}{2}$. The equation of the resulting curve is $r^2 = \cos 2\theta$ in polar coordinates and $(x^2+y^2)^2 = x^2 - y^2$ in Cartesian coordinates. Its shape is the familiar figure eight. (See Fig. 1.) A convenient reference for this material is the first chapter of Siegel's book **[8]**.

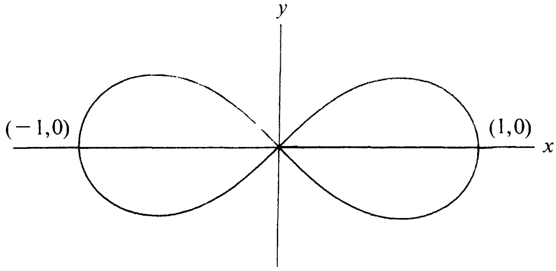


FIG. 1

If we use the formula for arc-length in polar coordinates $ds^2 = dr^2 + r^2 d\theta^2$, we find for the lemniscate $ds = dr/\sqrt{1-r^4}$. If we measure arc-length starting from the origin and passing into the first quadrant we see r increases from 0 to 1 and s is a monotonically increasing function of r. Explicitly, $s = \int_0^r dt/\sqrt{1-t^4}$. Let $\omega/2 = \int_0^1 dt/\sqrt{1-t^4}$. The total arc-length of the lemniscate is then 2ω. The constant ω is to the lemniscate what π is to the circle. To five places its value is 2.62057....

As s is a monotonically increasing function of r on $[0,1]$, r can be expressed as a function of s on $[0,\omega/2]$: set $r = \phi(s)$. This is Abel's notation. Gauss wrote sinlemn(s).

Abel shows that ϕ can be extended to a meromorphic function for a complex variable z. He shows that $\phi(z)$ is doubly periodic with

$$L = \langle 2\omega, 2\omega i\rangle = \{2m\omega + 2ni\omega \mid m,n \in Z\}$$

as a period lattice. The exact period lattice is $\langle (1+i)\omega, (1-i)\omega\rangle$. This fact will be useful later. He finds all the zeros and poles of $\phi(z)$. The zeros are the points of the lattice $\langle \omega, \omega i\rangle$ and the poles are obtained from the zeros by adding $(\omega + \omega i)/2$. Fig. 2 shows the location of the zeros and poles in a fundamental parallelogram $\{z \mid 0 \leqslant \operatorname{Re} z, \operatorname{Im} z < 2\omega\}$. He gives the following

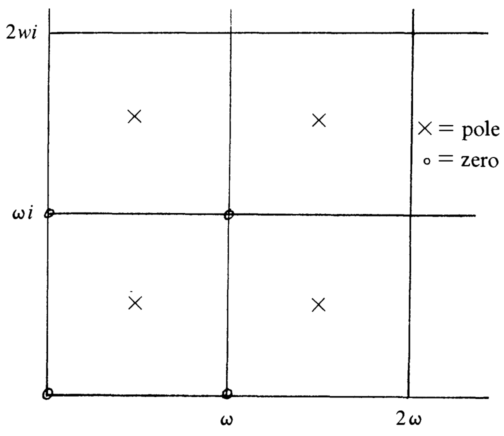


FIG. 2

product formula for $\phi(z)$:

$$\phi(z) = z\prod_{\alpha}\left(1 - \frac{z^4}{\alpha^4}\right)\prod_{\beta}\left(1 - \frac{z^4}{\beta^4}\right)^{-}$$

where α runs through the zeros and β through the poles in the region $0 \leqslant \arg z < \pi/2$. He also proves the addition formula (discovered much earlier by Euler; see **[8]**)

$$\phi(s+t) = \frac{\phi(s)\sqrt{1-\phi(t)^4} + \phi(t)\sqrt{1-\phi(s)^4}}{1+\phi(s)^2\phi(t)^2}.$$

All this and much else was discovered years earlier by Gauss, but had remained unpublished.

We will not prove these results. With the use of the modern theory of complex variables they may be considered exercises, albeit hard ones. It is quite amazing that Gauss, Abel, and Jacobi were able to prove all this when the theory of functions of a complex variable was still in its infancy.

To investigate the question of dividing the lemniscate into n equal parts with ruler and compass, we are reduced to asking the following question.

Question: For which integers n are the numbers $\phi(k2\omega/n)$, $k = 0, 1, \ldots, n-1$ constructible?

Note that the corresponding problem for the circle concerns the numbers $\sin(k2\pi/n)$.

3. Both for technical reasons and because the modern reader is much more likely to be familiar with the Weierstrass $\wp$ function than with Abel's function $\phi(z)$, we will reformulate the question of the previous section. Before doing so we briefly review the properties of the $\wp$ function which we will need. For proofs the reader can consult **[5]**, **[7]**, or **[8]**.

Let $\omega_1, \omega_2 \in \mathbb{C}$ be complex numbers such that ω_2/ω_1 is not real, and $\Lambda = \langle \omega_1, \omega_2 \rangle = \{m\omega_1 + n\omega_2 | m, n \in Z\}$, the corresponding lattice. A meromorphic function $f(z)$ on $\mathbb{C}$ is called elliptic with respect to Λ if $f(z + \lambda) = f(z)$ for all $z \in \mathbb{C}$ and $\lambda \in \Lambda$. The elements of Λ are called periods of $f(z)$. The set of functions which are elliptic with respect to Λ form a field which we denote by $\mathcal{M}(\Lambda)$. From another point of view, $\mathcal{M}(\Lambda)$ is the field of meromorphic functions on the Riemann surface $\mathbb{C}/\Lambda$.

Let $D(\Lambda) = \{r\omega_1 + s\omega_2 | 0 \leqslant r, s < 1\}$. $D(\Lambda)$ is called a fundamental parallelogram for Λ since the translates of $D(\Lambda)$ by elements of Λ simply cover the plane. If $f(z) \in \mathcal{M}(\Lambda)$ and $f(z)$ has no pole on $D(\Lambda)$, then $f(z)$ is a constant. This fundamental fact is proved as follows. If $f(z)$ has no pole on $D(\Lambda)$, it has no pole on $\mathbb{C}$ by periodicity. Thus $f(z)$ is continuous on the closure of $D(\Lambda)$, which is compact. It follows that $f(z)$ is bounded on $D(\Lambda)$ and by periodicity on all of $\mathbb{C}$. By Liouville's theorem a bounded entire function is a constant.

Does $\mathcal{M}(\Lambda)$ have any nonconstant functions? The answer is yes. Define

$$\wp(z;\Lambda) = z^{-2} + \Sigma'\left((z-\lambda)^{-2} - \lambda^{-2}\right)$$

where the sum is over all $\lambda \in \Lambda$, $\lambda \neq 0$. Since the lattice Λ is fixed in this discussion, we suppress it in the notation and write simply $\wp(z)$. Note that $\wp(-z) = \wp(z)$ and $\wp'(-z) = -\wp'(z)$; i.e., $\wp(z)$ is an even and $\wp'(z)$ is an odd function.

Both $\wp(z)$ and $\wp'(z)$ are in $\mathcal{M}(\Lambda)$, and in fact $\mathcal{M}(\Lambda) = \mathbb{C}(\wp(z), \wp'(z))$; i.e., every elliptic function with respect to Λ is a rational function of $\wp(z)$ and $\wp'(z)$. Moreover, $\wp(z)$ and $\wp'(z)$ are connected by the equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda)$$

where $g_2(\Lambda) = 60\Sigma'\lambda^{-4}$ and $g_3(\Lambda) = 140\Sigma'\lambda^{-6}$.

For $z \in \mathbb{C}$ let (z) be the unique element of $D(\Lambda)$ such that $z - (z) \in \Lambda$. The poles of $\wp(z)$ and $\wp'(z)$ are precisely at the points of Λ, i.e., those z such that $(z) = 0$. Those of $\wp(z)$ have

multiplicity 2 and those of $\mathcal{P}'(z)$ have multiplicity 3. It is not known where the zeros of $\mathcal{P}(z)$ lie but, if $a \notin \Lambda$, then the zeros of $\mathcal{P}(z) - \mathcal{P}(a)$ are precisely $\{z \in \mathbf{C} | (z) = (a) \text{ or } (z) = (-a)\}$. These are simple zeros unless $(a) = (-a)$, in which case they are double zeros. The zeros of $\mathcal{P}'(z)$ are $\{z \in \mathbf{C} | (z) = \omega_1/2, \omega_2/2, \text{ or } (\omega_1 + \omega_2)/2\}$. These are all simple zeros. These facts are often sufficient to enable us to write a function in $\mathfrak{M}(\Lambda)$ explicitly as a rational function of $\mathcal{P}$ and $\mathcal{P}'$.

A very important property of $\mathcal{P}$ and $\mathcal{P}'$ is the existence of an addition formula; i.e., for $z_1, z_2 \in \mathbf{C}$ with $z_1, z_2, z_1 + z_2 \notin \Lambda$, both $\mathcal{P}(z_1 + z_2)$ and $\mathcal{P}'(z_1 + z_2)$ can be expressed rationally in terms of $\mathcal{P}(z_1)$, $\mathcal{P}(z_2)$, $\mathcal{P}'(z_1)$, and $\mathcal{P}'(z_2)$. This can be seen from the following considerations. Let E be the complex points on the curve $y^2 = 4x^3 - g_2(\Lambda)x - g_3(\Lambda)$ together with a symbol ∞. This symbol represents the point at infinity on this curve. We call E the elliptic curve corresponding to Λ. Let ξ map $\mathbf{C}$ to E by $\xi(z) = (\mathcal{P}(z), \mathcal{P}'(z))$ for $z \notin \Lambda$ and $\xi(z) = \infty$ for $z \in \Lambda$. It can be shown that $\xi(z_1)$, $\xi(z_2)$, and $\xi(-z_1 - z_2)$ lie on a straight line. From this one deduces for $z_1, z_2 \notin \Lambda$ and $\mathcal{P}(z_1) \neq \mathcal{P}(z_2)$, i.e., $(z_1) \neq (\pm z_2)$, that

$$\mathcal{P}(z_1 + z_2) = -\mathcal{P}(z_1) - \mathcal{P}(z_2) + \frac{1}{4}\left(\frac{\mathcal{P}'(z_1) - \mathcal{P}'(z_2)}{\mathcal{P}(z_1) - \mathcal{P}(z_2)}\right)^2$$

and, if $(z_1) = (z_2) = z$, then

$$\mathcal{P}(2z) = -2\mathcal{P}(z) + \frac{1}{4}\left(\frac{\mathcal{P}''(z)}{\mathcal{P}'(z)}\right)^2.$$

If $z_1 + z_2 \in \Lambda$, i.e., $(z_1) = (-z_2)$, then the first formula continues to hold in the sense that both sides are infinite. The rational addition formula for $\mathcal{P}'(z)$ follows easily. We note that the coefficients of the rational functions giving the addition formula lie in the field generated over $\mathbf{Q}$ by $g_2(\Lambda)$ and $g_3(\Lambda)$.

It will be useful to give an algebraic interpretation of these results. By periodicity the map ξ gives rise to a map from $\mathbf{C}/\Lambda$ to E which we continue to call ξ. This map can be shown to be a bijection between $\mathbf{C}/\Lambda$ and E, and so by "transport of structure" E becomes a group. The group law on E is algebraic in the sense that there are rational functions f and g such that

$$(a,b) + (c,d) = (f(a,b,c,d), g(a,b,c,d)).$$

To see this let $\xi(z_1) = (a,b)$ and $\xi(z_2) = (c,d)$. By definition, $\xi(z_1) + \xi(z_2) = \xi(z_1 + z_2)$, or

$$(\mathcal{P}(z_1), \mathcal{P}'(z_1)) + (\mathcal{P}(z_2), \mathcal{P}'(z_2)) = (\mathcal{P}(z_1 + z_2), \mathcal{P}'(z_1 + z_2)).$$

It is now clear how the addition formulae give rise to the functions f and g.

Note that the identity element of the group E is $\xi(0) = \infty$. Let E_n be the points of order dividing n on E. Since $E \approx \mathbf{C}/\Lambda$, $E_n \approx \frac{1}{n}\Lambda/\Lambda \approx \Lambda/n\Lambda$. Thus E_n has n^2 elements. Explicitly,

$$E_n = \left\{\mathcal{P}\left(\frac{a\omega_1 + b\omega_2}{n}\right), \mathcal{P}'\left(\frac{a\omega_1 + b\omega_2}{n}\right) \middle| 0 \leqslant a, b < n\right\}.$$

For example, the points in E_2 are $\{\infty, (e_1, 0), (e_2, 0), (e_3, 0)\}$ where

$$e_1 = \mathcal{P}\left(\frac{\omega_1}{2}\right), \quad e_2 = \mathcal{P}\left(\frac{\omega_2}{2}\right), \quad \text{and} \quad e_3 = \mathcal{P}\left(\frac{\omega_1 + \omega_2}{2}\right).$$

Note that e_1, e_2, and e_3 are the roots of $4x^3 - g_2(\Lambda)x - g_3(\Lambda) = 0$. A consequence which is not obvious a priori is that this polynomial must have distinct roots.

4. We now return to a consideration of the lattice $L = \langle 2\omega, 2\omega i\rangle$. Let $\mathcal{P}(z)$ be the corresponding $\mathcal{P}$ function.

In the appendix we show that $g_2(L) = \frac{1}{4}$ and $g_3(L) = 0$. Thus, the elliptic curve corresponding to L is $y^2 = 4x^3 - \frac{1}{4}x$.

Notice that $iL = L$. From the definition of $\wp(z)$ we see that $\wp(iz) = -\wp(z)$ and by differentiation that $\wp'(iz) = i\wp(z)$. Using the addition formula we find after some calculation

$$\wp((1+i)z) = -\frac{i}{8}\frac{4\wp(z)^2 - \frac{1}{4}}{\wp(z)} \tag{1}$$

$$\wp((1-i)z) = \frac{i}{8}\frac{4\wp(z)^2 - \frac{1}{4}}{\wp(z)} \tag{2}$$

LEMMA 1. *If $\wp(\alpha)$ is constructible, so is $\wp(\alpha/2)$.*

Proof. In equation (1) substitute $z = \alpha/(1+i)$. We then see that $\wp(\alpha/(1+i))$ satisfies a quadratic equation with constructible coefficients. Thus $\wp(\alpha/(1+i))$ is constructible. In equation (2) substitute $z = \alpha/2$. Then, since $\wp((1-i)(\alpha/2)) = \wp(\alpha/(1+i))$ is constructible, $\wp(\alpha/2)$ satisfies a quadratic equation with constructible coefficients, and so $\wp(\alpha/2)$ is constructible.

COROLLARY. *Suppose $a,b,n \in Z$ with $n \geqslant 1$ and $ab \neq 0$. Then the numbers*

$$\wp((2a\omega + 2bi\omega)/2^n)$$

are constructible.

Proof. The numbers $\{\wp(\omega), \wp(i\omega), \wp((1+i)\omega)\}$ are the roots of $4x^3 - \frac{1}{4}x = 0$, i.e., $\{\frac{1}{4}, -\frac{1}{4}, 0\}$. This proves the result for $n = 1$. For general n the result follows by induction using Lemma 1.

We are now in a position to relate $\wp(z)$ and $\phi(z)$. Before doing so we make the simple remark that if $\wp(\alpha)$ is constructible so is $\wp'(\alpha)$ since $\wp'(\alpha)^2 = 4\wp(\alpha)^3 - \frac{1}{4}\wp(\alpha)$.

PROPOSITION 1. *$\phi(\alpha)$ is constructible if and only if $\wp(\alpha)$ is constructible.*

Proof. For the proof of Abel's theorem we only need the "if" part of the Proposition. We do this implication first.

The zeros and poles of $\phi(z)$ on $D(L)$ are $\{0, \omega, i\omega, (1+i)\omega\}$ and

$$\{(1+i)\omega/2, (3\omega + i\omega)/2, (\omega + 3i\omega)/2, (3\omega + 3i\omega)/2\},$$

respectively. The function

$$g(z) = \frac{\wp'(z)}{(\wp(z) - \wp(z_0))(\wp(z) - \wp(z_1))}$$

has the same zeros and poles if we set $z_0 = (1+i)\omega/2$ and $z_1 = (3\omega + i\omega)/2$. Thus $\phi(z) = Ag(z)$ for some constant A. Since $\phi(\omega/2) = 1$ and $g(\omega/2)$ is constructible by the corollary to Lemma 1, we see that A is constructible. If $\wp(\alpha)$ is constructible so is $g(\alpha)$ and thus so is $\phi(\alpha)$.

The proof of the converse is a bit more difficult. As we mentioned earlier $\phi(z)$ is periodic with respect to the lattice $M = \langle (1+i)\omega, (1-i)\omega \rangle$. Let $\wp_1(z)$ be the $\wp$ function corresponding to M. Since $(1+i)M = L$, we see (from definitions) that $2i\wp((1+i)z) = \wp_1(z)$.

The zeros and poles of $\phi(z)$ on $D(M)$ are $\{0, \omega\}$ and $\{(1+i)\omega/2, (1-i)\omega/2\}$, respectively. Comparing zeros and poles we see there is a constant B such that

$$\phi(z) = B\frac{\wp_1(z) - \wp_1(\omega)}{\wp_1'(z)}.$$

Evaluating at $z = \omega/2$ and using the corollary to Lemma 1 once again we see that B is constructible.

Let $u_0 = (1+i)\omega/2$ and $u_1 = (1-i)\omega/2$. Since the zeros of $\mathcal{P}_1'(z)$ are u_0, u_1, and ω we find

$$\phi(z)^2 = \frac{B^2}{4}\frac{\mathcal{P}_1(z) - \mathcal{P}_1(\omega)}{(\mathcal{P}_1(z) - \mathcal{P}_1(u_0))(\mathcal{P}_1(z) - \mathcal{P}_1(u_1))}.$$

It follows that if $\phi(\alpha)$ is constructible so is $\mathcal{P}_1(\alpha)$.
From previous results we have

$$\mathcal{P}_1(z) = 2i\mathcal{P}((1+i)z) = \frac{1}{4}\frac{4\mathcal{P}(z)^2 - \frac{1}{4}}{\mathcal{P}(z)}.$$

Thus, if $\mathcal{P}_1(\alpha)$ is constructible so is $\mathcal{P}(\alpha)$. This completes the proof.

In Section 2 we showed that the lemniscate can be divided into n equal parts with ruler and compass if and only if the numbers $\{\phi(k2\omega/n) | k = 1,2,\ldots,n-1\}$ are constructible. In the light of Proposition 1 we are reduced to the question of finding those integers n such that the numbers $\{\mathcal{P}(k2\omega/n) | k = 1,2,\ldots,n-1\}$ are constructible.

5. Recall that E represents the complex points on the elliptic curve $y^2 = 4x^3 - \frac{1}{4}x$ together with the point at infinity. By means of $\xi(z) = (\mathcal{P}(z), \mathcal{P}'(z))$ we have an isomorphism between $\mathbf{C}/L$ and E. We now argue with E the way we did with C in our "elaborate" proof of Gauss's theorem. Since E_n is finite, we see by the same reasoning that we applied to C_n in Section 1 that the coordinates of the points in E_n are algebraic over $\mathbf{Q}$. Adjoin these coordinates to $\mathbf{Q}$ and call the resulting field K_n. Then $K_n/\mathbf{Q}$ is Galois. Let G_n denote its Galois group. G_n acts on E_n and gives rise to a monomorphism from G_n into

$$\mathrm{Aut}(E_n) \approx \mathrm{Aut}(L/nL) \approx \mathrm{Aut}(Z/nZ \oplus Z/nZ) \approx \mathrm{Gl}_2(Z/nZ).$$

At this point we do not know much about the image of G_n. Moreover, the order of $\mathrm{Gl}_2(Z/nZ)$ is never a power of two. We seem to have reached a dead end.

The situation is saved by the realization that there is some additional structure which has not been used, namely, $L = \langle 2\omega, 2\omega i\rangle = Z[i](2\omega)$; i.e., L is a $Z[i]$ module of rank one, not just an abelian group. In general, a lattice Λ is said to admit complex multiplication if the ring $\{\alpha \in \mathbf{C} | \alpha L \subseteq L\}$ is properly bigger than Z.

Since L is a $Z[i]$ module, so is $\mathbf{C}/L$ and via ξ we can make E into a $Z[i]$ module. Since, as we have seen, $\mathcal{P}(iz) = -\mathcal{P}(z)$ and $\mathcal{P}'(iz) = i\mathcal{P}'(z)$, the action of i on E is given by $i(x,y) = (-x, iy)$.

LEMMA 2. *$E_n \approx Z[i]/nZ[i]$ as $Z[i]$ modules.*

Proof. $E_n \approx \frac{1}{n}L/L \approx L/nL \approx Z[i]/nZ[i]$.

Let $F = \mathbf{Q}(i)$ and adjoin the coordinates of E_n to F. Call the resulting field F_n and let $\mathcal{G}_n$ be its Galois Group over F. Since $\mathcal{G}_n$ leaves i fixed, the action of $\mathcal{G}_n$ on E_n preserves the $Z[i]$ module structure. Thus we get a monomorphism from $\mathcal{G}_n$ into the $Z[i]$ automorphisms of E_n. Now, by Lemma 2 ,

$$\mathrm{Aut}_{Z[i]}(E_n) \approx \mathrm{Aut}_{Z[i]}(Z[i]/nZ[i]) \approx (Z[i]/nZ[i])^*.$$

We have shown

PROPOSITION 2. *The group $\mathcal{G}_n$ is abelian. If $(Z[i]/nZ[i])^*$ is a two-group, then the numbers $\mathcal{P}((2a\omega + 2bi\omega)/n)$ and $\mathcal{P}'(2a\omega + 2bi\omega)/n)$ are constructible.*

Abel's Theorem now follows from the following easily proved Lemma.

LEMMA 3. *$(Z[i]/nZ[i])^*$ is a two-group if and only if n is a power of 2 times a product of distinct Fermat primes.*

We now turn to the proof of the converse to Abel's theorem.

LEMMA 4. *Let M be the field generated over $F = \mathbf{Q}(i)$ by adjoining $\wp(2\omega/n)^2$. Then M/F is Galois and the Galois group is isomorphic to $(Z[i]/nZ[i])^*$ modulo the image of the group $\{\pm 1, \pm i\}$.*

Proof. Let $L_0 = Z[i]$ be considered as a lattice in $\mathbf{C}$, and $\wp_0(z)$ the corresponding $\wp$ function. Let $h(z) = g_2(L_0)^{-1}\wp_0(z)^2$. It follows from the arithmetic copy of complex multiplication that $F(h(1/n))$ is the ray class field of F corresponding to the modulus n (see page 135 of [5]). The ray class group of modulus n is precisely the group described in the statement of the lemma. We will show $h(1/n) = 4\wp(2\omega/n)^2$ and that will complete the proof.

From the definition of $\wp$ and $\wp_0$ we see easily that $\wp(2\omega z) = (2\omega)^{-2}\wp_0(z)$. In the appendix we will show $\Sigma'\gamma^{-4} = \omega^4/15$ where the sum is over all nonzero elements of $Z[i]$. Thus $g_2(L_0) = 60\Sigma'\gamma^{-4} = 4\omega^4$. It follows that $h(z) = (4\omega^4)^{-1}(2\omega)^4\wp(2\omega z)^2 = 4\wp(2\omega z)^2$.

THEOREM 2. *If the lemniscate can be divided into n equal parts with ruler and compass, then n is a power of two times a product of distinct Fermat primes.*

Proof. If the hypothesis holds then $\phi(2\omega/n)$ is a constructible number. By Proposition 1, $\wp(2\omega/n)$ is constructible. It then follows from Lemma 4 that $(Z[i]/nZ[i])^*$ is a two-group. The result is now a consequence of the "only if" part of Lemma 3.

Appendix. In Section 3 we asserted that the pair of functions $\wp$ and $\wp'$ corresponding to the lattice $L = \langle 2\omega, 2\omega i\rangle$ parametrize the elliptic curve $y^2 = 4x^3 - \frac{1}{4}x$. To prove this we proceed as follows.

As we have seen, $\wp$ and $\wp'$ parametrize $y^2 = 4x^3 - g_2(L)x - g_3(L)$ where

$$g_2(L) = 60\sum{}'\gamma^{-4} \quad \text{and} \quad g_3(L) = 140\sum{}'\gamma^{-6}.$$

Since $iL = L$ and $i^6 = -1$ we see $g_3(L) = 0$. To show $g_2(L) = \frac{1}{4}$ it is equivalent to show

PROPOSITION 3. $\Sigma(r + si)^{-4} = \omega^2/15$ *where the sum is over all nonzero Gaussian integers.*

This result was obtained by Hurwitz in [4]. In fact, he shows that, more generally, $\Sigma(r + si)^{-4n} = ((2\omega)^{4n}/(4n)!)E_n$ where the E_n are positive rational numbers. Hurwitz shows that these rational numbers have many properties analogous to the Bernoulli numbers, including an analogue of the von Staudt–Clausen theorem. Nowadays these numbers E_n are called Hurwitz numbers in his honor. It is worth noting that Gauss, in the 61st entry in his mathematical diary (see pages 515 and 516 of [3]), states a result which is equivalent to the assertion that the E_n are rational.

Hurwitz's proof of the rationality of the E_n is quite easy. However, we prefer to give another proof which has the flavor of Euler's proof that $\Sigma n^{-2} = \pi^2/6$ and which depends only on the product formula for $\phi(z)$.

For a lattice L in $\mathbf{C}$ let $|L| = \Sigma'\gamma^{-4}$, the sum being over $L - \{0\}$. Consider the three lattices $L_0 = \langle\omega, \omega i\rangle$, $L_1 = \{(m + ni/2)\omega \mid m \text{ and } n \text{ odd}\}$ and $L_2 = \{(m + ni/2)\omega \mid m \text{ and } n \text{ of opposite parity}\}$. Then $\frac{1}{2}L_0 = L_1 \cup L_2 \cup L_0$ where the union is disjoint. Clearly $|\frac{1}{2}L_0| = 16|L_0|$. It is easily checked that $(1 + i/2)L_1 = L_2$ and it follows that $|L_2| = (2/(1+i))^4|L_1| = -4|L_1|$. Putting all this together we have $|L_1| = -5|L_0|$.

As shown by Gauss and Abel

$$\phi(z) = z\prod_{\alpha}\left(1 - \frac{z^4}{\alpha^4}\right)\prod_{\beta}\left(1 - \frac{z^4}{\beta^4}\right)^{-1}$$

where $\alpha \in L_0$, $\beta \in L_1$, and $0 \leqslant \arg\alpha, \arg\beta < \pi/2$. Taking the logarithmic derivative of both

sides yields

$$z\frac{\phi'(z)}{\phi(z)} = 1 + (|L_1| - |L_0|)z^4 + \cdots.$$

We have to evaluate the left-hand side in a different way. From $z = \int_0^{\phi(z)} dt/\sqrt{1-t^4}$ we find $\phi'(z)^2 = 1 - \phi(z)^4$. Let $\phi(z) = z + cz^5 + \cdots$ be the power series expansion of $\phi(z)$ about $z = 0$. Substituting in $\phi'(z)^2 = 1 - \phi(z)^4$ and comparing coefficients of z^4 we find $c = -\frac{1}{10}$. From this we derive $z(\phi'(z)/\phi(z) = 1 - \frac{2}{5}z^4 + \cdots$.

Thus $|L_1| - |L_0| = -\frac{2}{5}$. Since also $|L_1| = -5|L_0|$, it follows that $|L_0| = \frac{1}{15}$; i.e.,

$$\sum (r + si)^{-4} = \frac{\omega^4}{15}.$$

From Proposition 3 we see that the first Hurwitz number E_1 is equal to $\frac{1}{10}$. It is now relatively simple to show that E_n is rational for all $n \geqslant 1$. For this purpose we sketch the proof of a more general proposition.

For a lattice Λ define $s_m(\Lambda) = \Sigma'\lambda^{-m}$, where $m > 2$ is an integer and the sum is over all $\lambda \in \Lambda, \lambda \neq 0$. These sums are convergent. Since $\Lambda = -\Lambda$, it follows that $s_m(\Lambda) = 0$ for m odd.

Let $\wp(z)$ be the $\wp$ function corresponding to the lattice Λ and $\wp(z) = 1/z^2 + \Sigma_{n=1}^{\infty} b_n z^{2n}$ the Laurent series of $\wp(z)$ about $z = 0$. From the definition one can calculate that $b_n = (2n+1)s_{2n+2}(\Lambda)$. See page 10 of [5] for details.

PROPOSITION 4. *Let* $\mathcal{O} = \{\gamma \in \mathbb{C} | \gamma\Lambda \subseteq \Lambda\}$ *and suppose* $\Lambda = \mathcal{O}\omega$ *for some* $\omega \in \mathbb{C}$. *Suppose further that* $s_4(\Lambda)$ *and* $s_6(\Lambda)$ *are in* $\mathbf{Q}$. *Then* $s_m(\Lambda) \in \mathbf{Q}$ *for all* $m \geqslant 4$ *and*

$$\sum_{\gamma \in \mathcal{O}}' \gamma^{-2n} = s_{2n}(\Lambda)\omega^{2n}.$$

Proof. Differentiating the fundamental relation $\wp'(z)^2 = 4\wp(z)^3 - g_2(\Lambda)\wp(z) - g_3(\Lambda)$ we find $\wp''(z) = 6\wp(z)^2 - \frac{1}{2}g_2(\Lambda)$. Equating coefficients leads to the following recursion formula for b_n,

$$(2n+3)(n-2)b_n = 3\sum_{k=1}^{n-2} b_k b_{n-k-1}.$$

This in turn shows that

$$s_{2n+2}(\Lambda) = \sum_{k=1}^{n-2} \gamma_{n,k} s_{2k+2}(\Lambda) s_{2n-2k}(\Lambda)$$

where $\gamma_{n,k} \in \mathbf{Q}$. Thus, if $s_4(\Lambda)$ and $s_6(\Lambda)$ are in $\mathbf{Q}$ it follows by induction that $s_m(\Lambda) \in \mathbf{Q}$ for all $m \geqslant 4$. Since every element of Λ is uniquely of the form $\gamma\omega$ with $\gamma \in \mathcal{O}$, the proposition follows immediately.

For our lattice $L = \langle 2\omega, 2\omega i\rangle$ we have $\mathcal{O} = Z[i]$, $s_4(L) = \frac{1}{15}$ (by Proposition 3), and $s_6(L) = 0$. The rationality of E_n for all $n \geqslant 1$ follows from Proposition 4.

References

1. N. H. Abel, Oeuvres Complètes, Nouvelle Edition, Oslo, 1881.
2. C. F. Gauss, Disquisitiones Arithmeticae, transl. by Arthur A. Clarke, Yale University Press, New Haven, 1966.
3. ______, Carl Friedrich Gauss Werke, vol. 10, Royal Scientific Society of Göttingen, 1917.
4. A. Hurwitz, Über die Entwicklungskoeffizienten der lemniscatischen Funktionen, Mathematische Werke, vol. 2, pp. 342–373, Birkhäuser Verlag, Basel und Stuttgart, 1962.
5. S. Lang, Elliptic Functions, Addison-Wesley, Reading, Mass., 1973.
6. O. Ore, Niels Henrik Abel, Mathematician Extraordinary, University of Minnesota Press, Minneapolis, 1957.
7. A. Robert, Elliptic Curves, Springer Lecture Notes 326, Springer-Verlag, New York, 1973.
8. C. L. Siegel, Topics in Complex Function Theory, vol. 1, Wiley-Interscience, New York, 1969.

ITERATING ANALYTIC SELF-MAPS OF DISCS

R. B. BURCKEL

Department of Mathematics, Kansas State University, Manhattan, KS 66506

Introduction. Expressions of the form $x^{x^{x}}$ were considered already by Leonhard Euler. What does this expression mean? If a limit is involved, for what x does it exist and what is its value? For positive real x the question of meaning is easy: we set $x_1 = x$ and inductively $x_n = x^{x_{n-1}}$ for $n > 1$. The convergence question here is subtle but not too difficult. (Surprisingly, the sequence $\{x_n\}$ converges for some $x < 1$ and also for some $x > 1$.) For a discussion of this particular problem, with full historical references, see Shell [**18**], Andrews and Lacher [**1**], and Knoebel [**12**]. What we have here is some sort of indefinite iteration of an exponential function, and the problems become stickier if x is allowed to be complex. For example, what sense is to be made of $i^{i^{i}}$?

This point of view of iterating a function is what will be considered here. For a rather large class of functions we can answer the question of convergence of the sequence of iterates, and sometimes we can even find explicitly the value of the limit; e.g., the theory of section 5 below shows that $i^{i^{i}}$ converges but does not give its value. On the other hand (since a half-plane is conformal to a disc), the reader will have no trouble applying Theorem 5.3 to infer that the sequence of iterates of the function $f(z) = az + (1-a)/z$, where $0 < a < 1$, converges to 1 when $\operatorname{Re} z > 0$ and converges to -1 when $\operatorname{Re} z < 0$. (This latter problem is treated directly, and the case $\operatorname{Re} z = 0$ is also settled in exercise III.262 of Pólya and Szegő [**16**].)

There is a large literature on the subject of iterating holomorphic (especially entire or rational) functions, the work of Gaston Julia, Pierre Fatou, and Irving Baker being among the most important. In this paper we look at functions f defined and holomorphic in $D = \{z \in \mathbb{C} : |z| < 1\}$ with $f(D) \subset D$. Because the range of f lies in its domain, iteration of f is possible. That is, we set $f^{[0]} = I$, the identity function on D, $f^{[1]} = f$ and generally $f^{[n+1]} = f \circ f^{[n]}$ for any positive integer n, and we study the sequence $\{f^{[n]}\}$ of *iterates* of f. This can be done for $\mathbb{C}$ or any region Ω in the role of D, as long as $f(\Omega) \subset \Omega$. For the disc the results are quite complete and satisfying, and this is not as special a case as it appears. For if $\phi : D \to \Omega$ is a *conformal* (= one-to-one, onto, holomorphic) map with a continuous extension to $\bar{D}$, then consideration of $F = \phi^{-1} \circ f \circ \phi$ reduces the case Ω to the case D, since evidently $f^{[n]} = \phi \circ F^{[n]} \circ \phi^{-1}$ transmits to the sequence $\{f^{[n]}\}$ on Ω whatever properties the sequence $\{F^{[n]}\}$ has on D. Thanks to the Osgood-Taylor-Carathéodory extension of the Riemann Mapping Theorem, such a ϕ is available whenever Ω is the interior of a Jordan curve. So the disc situation is really rather general.

There is a close connection between convergence of the sequence $\{f^{[n]}\}$ and the existence of fixed points of f in D. Here and in the sequel, *convergence* for a sequence of holomorphic functions means *local uniform convergence*, i.e., uniform convergence in a neighborhood of each point of D, or, equivalently, uniform convergence on each compact subset of D. A point $z_0 \in D$ is a *fixed point* of f if $f(z_0) = z_0$. If $f^{[n]} \to F$, then

$$F(z) = \lim_n f^{[n+1]}(z) = \lim_n f\big(f^{[n]}(z)\big) = f\Big(\lim_n f^{[n]}(z)\Big) = f(F(z))$$

for every $z \in D$ such that $F(z) \in D$, by continuity of f at $F(z)$. Thus each such $F(z)$ is a fixed point of f. But it is an easy consequence of Schwarz's Lemma (see below) that no holomorphic function in D except the identity fixes more than one point, and so, for f not the identity function,

The author received his Ph.D. from Yale under the direction of Frank Hahn. He has taught at the University of Oregon and now holds a professorship at Kansas State University. During 1977–78 he was Gastprofessor at the Universität des Saarlandes. His interests are in classical and functional analysis. He has published books on Weakly Almost Periodic Functions on Semigroups, Characterizations of $C(X)$ Among its Subalgebras, An Introduction to Classical Complex Analysis, and has translated Heinz Bauer's Probability Theory into English.—*Editors*

F is constant. Now fixed points often exist; for example, L. E. J. Brouwer's famous theorem asserts that every *continuous* $f: \overline{D} \to \overline{D}$ possesses at least one fixed point in $\overline{D}$. Moreover, classical "normality" criteria, to be discussed below, ensure that $\{f^{[n]}\}$ has many convergent subsequences. In fact, classical normality is modern precompactness, and the set $\{f^{[n]}\}$ is nowadays thought of as a semigroup. Thus a compact, commutative semigroup is on the scene. There is a well-worked-out, even rather elementary, theory of such animals ready at hand. Hence we have all the ingredients for an interesting theory and also some cause for optimism that positive results may be forthcoming.

1. Prerequisites and Tools. Our needs from function theory will be quite modest indeed, facts surely known to most readers and contained in every elementary text on the subject. We list them for easy reference later and emphasize that nothing beyond these basic facts is needed.

1.1. UNIQUENESS THEOREM. *If f and g are holomorphic in the open, connected set* (= region)Ω *and $f = g$ in some nonempty open subset of Ω, then $f = g$ throughout Ω.*

1.2. OPEN MAP THEOREM. *The image of a region under any nonconstant holomorphic function is an open set.*

1.3. ROUCHÉ'S THEOREM. *If f and g are holomorphic in a neighborhood of $\overline{D}(c,r) = \{z \in \mathbb{C} : |z - c| \leqslant r\}$ and $|f - g| < |g|$ on $\partial D(c,r)$, then f has as many zeros in $D(c,r)$ as does g.*

1.4. SCHWARZ'S LEMMA. *If $f: D \to D$ is holomorphic and $f(0) = 0$, then $|f(z)| \leqslant |z|$ for all $z \in D$.*

1.5. WEIERSTRASS'S THEOREM. *A locally uniform limit of holomorphic functions is a holomorphic function.*

1.6. MONTEL'S THEOREM. *Every uniformly bounded family of holomorphic functions in a region is a* normal family; *that is, any sequence in it contains a subsequence which is locally uniformly convergent throughout the region.* (*The limit is holomorphic by* 1.5.)

1.7. COROLLARY TO MONTEL'S THEOREM. *If $\{f_n\}$ is a uniformly bounded sequence of holomorphic functions in a region Ω and if every convergent subsequence of $\{f_n\}$ has the same limit, then the sequence $\{f_n\}$ is convergent.*

Proof. Let the common subsequential limit be g and suppose $\{f_n\}$ does not converge locally uniformly in Ω to g. This means that for some compact $K \subset \Omega$, some $\varepsilon > 0$ and some $n_1 < n_2 < \cdots$ we have $\sup_K |f_{n_j} - g| \geqslant \varepsilon$ for every j. However, $\{f_{n_j}\}$ contains a locally uniformly convergent subsequence by 1.6. By hypothesis its limit is g, yet its convergence to g is not uniform on K, contradiction.

2. Conformal Automorphisms of Discs. For each $a \in D$ let T_a denote the rational function (so-called *fractional linear* or *Möbius transformation*)

$$T_a(z) = \frac{z - a}{1 - \bar{a}z}.$$

It is defined for all $z \in \mathbb{C}$ if $a = 0$ and all $z \in \mathbb{C} \setminus \{1/\bar{a}\}$ otherwise. In any case it is always defined in a whole neighborhood of $\overline{D}$, and in what follows its domain will be considered to be $\overline{D}$. Let us review quickly the elementary theory of these functions.

2.1. THEOREM. *For each $a \in D$ the function T_a maps $\overline{D}$ univalently onto $\overline{D}$, carrying D onto D and $T_a^{-1} = T_{-a}$. Conversely, any conformal map of D onto D has the form uT_a for some $a \in D$, $u \in \partial D$.*

Proof. One calculates that for any $z \in \overline{D}, a \in D$,

$$|T_a(z)|^2 = 1 - \frac{(1 - |z|^2)(1 - |a|^2)}{|1 - \bar{a}z|^2}.$$

This shows at once that T_a maps $\bar{D}$ into $\bar{D}$, D into D, and ∂D into ∂D. Consequently, $T_a \circ T_{-a}$ and $T_{-a} \circ T_a$ may be formed. A simple calculation reveals that each equals the identity function on $\bar{D}$.

If F is a conformal map of D onto D, let $a = F^{-1}(0)$ and consider $f = F \circ T_a^{-1}$. This is again a conformal map of D onto D, by the result of the first paragraph. Since $f(0) = 0$, we can apply Schwarz's Lemma to both f and f^{-1} to get

$$|z| = |f^{-1}(f(z))| \leqslant |f(z)| \leqslant |z| \qquad \forall z \in D.$$

The holomorphic function $f(z)/z$ thus has constant modulus 1 and so is constant, by the Open Map Theorem. Calling this unimodular constant u, we have $f = uI$, so $F = f \circ T_a = uT_a$.

As a corollary we can formulate an "invariant" version of Schwarz's Lemma:

2.2. THEOREM. *If $F: D \to D$ is holomorphic, then*

$$\left| \frac{F(z) - F(w)}{1 - \bar{F}(w)F(z)} \right| \leqslant \left| \frac{z - w}{1 - \bar{w}z} \right| \qquad \forall z, w \in D.$$

Proof. Given $w \in D$, apply Schwarz's Lemma to $f = T_{F(w)} \circ F \circ T_w^{-1}$.

It is well known and easy to prove that each T_a maps every disc lying in D onto another. What we shall need is the following slightly more precise version of this: Write $D(c,r)$ for the open disc $\{z \in \mathbb{C} : |z - c| < r\}$, $c \in \mathbb{C}$, $r \geqslant 0$. Then:

2.3. THEOREM. *For each $a \in D$, $0 \leqslant r < 1$ the set $\{z \in \mathbb{C} : |T_a(z)| \leqslant r\} = T_a^{-1}(\bar{D}(0,r))$ is the closed disc with the center $C = (1 - r^2)a/(1 - |a|^2 r^2)$ and radius $R = (1 - |a|^2)r/(1 - |a|^2 r^2)$. It lies in D.*

Proof. By multiplying out everything, one sees first that

$$|T_a(z)|^2 \leqslant r^2 \Leftrightarrow (1 - |a|^2 r^2)|z|^2 + 2(r^2 - 1)\mathrm{Re}(\bar{a}z) \leqslant r^2 - |a|^2$$

and then, by "completing the square," that the latter inequality is equivalent to $|z - C|^2 \leqslant R^2$. Moreover, calculation shows that

$$|C| + R = 1 - \frac{(1 - |a|)(1 - r)}{1 + |a|r} < 1$$

and therefore $D(C,R)$ lies in D (a fact which also follows from 2.1).

An important and useful representation of T_a is provided by the next result.

2.4. THEOREM. *Let $a \in D$, $u \in \partial D$, $T = uT_a$ and $z_1, z_2 \in \bar{D}$. Then*

$$\frac{Tz - Tz_1}{Tz - Tz_2} = \frac{1 - \bar{a}z_2}{1 - \bar{a}z_1} \cdot \frac{z - z_1}{z - z_2} \qquad \forall z \in D \setminus \{z_2\}.$$

Proof. This is a straightforward algebraic simplification of the quotient on the left side of the equation.

If z_1, z_2 are distinct and are fixed points of T and if we write λ for $(1 - \bar{a}z_2)/(1 - \bar{a}z_1)$ and $S(w) = (w - z_1)/w - z_2)$ $(w \in \mathbb{C} \setminus \{z_2\})$, then the last theorem asserts that

$$S(T(z)) = \lambda S(z)$$

$$T(z) = S^{-1}(\lambda S(z)) \qquad \forall z \in D \setminus \{z_2\}.$$

This makes iteration of T particularly simple; for, as noted in the introduction, we then have

$$T^{[n]}(z) = S^{-1}(\lambda^n S(z)).$$

And we see too the importance of whether $|\lambda| < 1$ or $|\lambda| \geqslant 1$ to the question of convergence of

the sequence $\{T^{[n]}\}$ of iterates of T. The facts about the fixed points z_1, z_2 and the *multiplier* λ are summarized in the next two theorems.

2.5. THEOREM. *For each $a \in D$, $u \in \partial D$, the map uT_a of $\bar{D}$ onto $\bar{D}$ either is the identity map or has one or two fixed points. In the latter case the two fixed points lie on ∂D.*

Proof. Suppose $uT_a \neq I$. The statement $uT_a(z) = z$ for some $z \in \bar{D}$ is equivalent to

$$\bar{a}z^2 + (u-1)z - ua = 0. \tag{$*$}$$

If $a = 0$, then $u \neq 1$ (since $uT_a \neq I$) and there is exactly one $z \in \mathbb{C}$ which satisfies $(*)$, namely, $z = 0$. Now suppose $a \neq 0$. Then $z = 0$ is not a root of $(*)$ and, remembering that $\bar{u} = 1/u$, we see that for any $z \neq 0$

$$\bar{a}\left(\frac{1}{\bar{z}}\right)^2 + (u-1)\left(\frac{1}{\bar{z}}\right) - ua = -\frac{u}{\bar{z}^2}\left[\bar{a}z^2 + (u-1)z - ua\right]^- .$$

Consequently, $z \neq 0$ is a root of $(*)$ if and only if $1/\bar{z}$ is a root of $(*)$. Now the quadratic equation $(*)$ has one or two roots. In the latter case, if both lie in $\bar{D}$, then both in fact have modulus 1, since otherwise their reciprocal conjugates would lie outside $\bar{D}$ and consequently constitute two *additional* roots.

2.6. THEOREM. *If $a \in D$, $u \in \partial D$ and uT_a has exactly two fixed points z_1 and z_2 in $\bar{D}$, then the number $\lambda = (1 - \bar{a}z_2)/(1 - \bar{a}z_1)$ is not unimodular.*

Proof. For $j = 1,2$, the equation $z_j = uT_a(z_j) = u(z_j - a)/(1 - \bar{a}z_j)$ implies $1 - \bar{a}z_j = u(z_j - a)/z_j = u(1 - (a/z_j)) = u(1 - a\bar{z}_j)$, since $z_j\bar{z}_j = 1$ by 2.5. It follows at once that $\lambda = \bar{\lambda}$; so λ is real. Now $a \neq 0$ (else $uT_a = uI$ has one or infinitely many fixed points in $\bar{D}$). Since moreover $z_1 \neq z_2$, it is clear that $\lambda - 1 = \bar{a}(z_1 - z_2)/(1 - \bar{a}z_1) \neq 0$. Finally, $|\bar{a}(z_1 + z_2)| \leqslant |a|(|z_1| + |z_2|) = 2|a| < 2$, so $\lambda + 1 = [2 - \bar{a}(z_1 + z_2)]/(1 - \bar{a}z_1) \neq 0$.

REMARK. Though this fact is not needed in what follows, one can prove with only a little more effort that λ is positive.

In case there is only one fixed point on ∂D, there is again available a map S for which $S \circ T \circ S^{-1}$ takes a particularly simple form for iteration. The facts are all summarized in:

2.7. THEOREM. *If $a \in D$, $u \in \partial D$ and $T = uT_a$ has a unique fixed point z_0 on ∂D, then $u \neq -1$ and*

$$\frac{z_0}{Tz - z_0} = \frac{u-1}{u+1} + \frac{z_0}{z - z_0} \qquad \forall z \in D.$$

Proof. By 2.5, z_0 is the unique fixed point in $\bar{D}$. That is, z_0 is the unique root of the quadratic equation $(*)$ above. But $a \neq 0$ (else 0 is a fixed point) so z_0 is the unique root of

$$z^2 + \left(\frac{u-1}{\bar{a}}\right)z - \frac{ua}{\bar{a}} = 0.$$

Therefore

$$\frac{u-1}{\bar{a}} = -2z_0$$

$$a = \frac{(1-\bar{u})}{2}z_0.$$

In particular, since $|a| < 1 = |z_0|$, we see that $u \neq -1$. Using this value of a in the definition of T, the asserted identity is verified by a routine calculation.

3. Julius Wolff's Theorem. The following beautiful geometric result is the key to all the deeper facts about sequences of iterates.

3.1. THEOREM [25]. *Let $f: D \to D$ be holomorphic and fixed-point-free. Then there exists $u \in \partial D$ with the following property: every closed disc in D which is tangent to ∂D at u is mapped into itself by every iterate of f.*

Proof. Pick $r_k \in (0,1)$ with $r_k \uparrow 1$ and set $f_k = r_k f$. Since

$$|I - (I - f_k)| < r_k = |I| \text{ on } \partial D(0, r_k),$$

it follows from Rouché's Theorem that $I - f_k$ has as many zeros in $D(0, r_k)$ as does I. Thus $I - f_k$ has at least one zero in $D(0, r_k)$. Let a_k be such a zero. Thus

$$r_k f(a_k) = a_k \qquad \forall k. \tag{1}$$

Pass to a subsequence of $\{a_k\}$ and so assume with no loss of generality that

$$\lim_{k \to \infty} a_k = u \tag{2}$$

exists in $\bar{D}$. If $u \in D$, then letting $k \to \infty$ in (1) shows that $f(u) = u$. Since f is supposed to be fixed-point free, this is not possible. Hence

$$|u| = 1. \tag{3}$$

Since each $f_k^{[n]}$ maps D into D, 2.2 says that

$$\left| \frac{f_k^{[n]}(z) - f_k^{[n]}(a_k)}{1 - \overline{f_k^{[n]}(a_k)} f_k^{[n]}(z)} \right| \leqslant \left| \frac{z - a_k}{1 - \bar{a}_k z} \right| \qquad \forall z \in D, \quad k, n \in \mathbb{N}.$$

That is,

$$\left| \frac{f_k^{[n]}(z) - a_k}{1 - \bar{a}_k f_k^{[n]}(z)} \right| \leqslant \left| \frac{z - a_k}{1 - \bar{a}_k z} \right| \qquad \forall z \in D, \quad k, n \in \mathbb{N}. \tag{4}$$

If we set

$$r_k(z) = \left| \frac{z - a_k}{1 - \bar{a}_k z} \right| \in [0, 1), \tag{5}$$

$$C_k(z) = \frac{(1 - r_k^2(z)) a_k}{1 - |a_k|^2 r_k^2(z)}, \tag{6}$$

$$R_k(z) = \frac{(1 - |a_k|^2) r_k(z)}{1 - |a_k|^2 r_k^2(z)}, \tag{7}$$

then according to 2.3 the set

$$S_k(z) = \left\{ w \in \mathbb{C} : \left| \frac{w - a_k}{1 - \bar{a}_k w} \right| \leqslant r_k(z) \right\}$$

is the closed disc in D with center $C_k(z)$ and radius $R_k(z)$. According to (4) we have $z, f_k^{[n]}(z) \in S_k(z)$. That is,

$$|f_k^{[n]}(z) - C_k(z)| \leqslant R_k(z) \qquad \forall z \in D, \quad k \in \mathbb{N}, \quad n = 0, 1, 2, \ldots \tag{8}$$

Evidently z in fact lies on the boundary of this disc. Now a calculation shows that

$$1 - |a_k|^2 r_k^2(z) = \frac{(1 - |a_k|^2)[1 + |a_k|^2 - 2\mathrm{Re}(\bar{a}_k z)]}{|1 - \bar{a}_k z|^2} \tag{9}$$

and so

$$\frac{(|1 - |a_k|^2) r_k(z)}{1 - |a_k|^2 r_k^2(z)} = \frac{|1 - \bar{a}_k z|^2 r_k(z)}{1 + |a_k|^2 - 2\mathrm{Re}(\bar{a}_k z)}. \tag{10}$$

Another calculation shows that

$$1 - r_k^2(z) = \frac{(1 - |a_k|^2)(1 - |z|^2)}{|1 - \bar{a}_k z|^2},$$

and so, using (9),

$$\frac{(1 - r_k^2(z))a_k}{1 - |a_k|^2 r_k^2(z)} = \frac{(1 - |z|^2)a_k}{1 + |a_k|^2 - 2\mathrm{Re}(\bar{a}_k z)}. \tag{11}$$

Now it is clear from (5) and (2) that as $k \to \infty$

$$r_k(z) = \left|\frac{z - a_k}{1 - \bar{a}_k z}\right| \to \left|\frac{z - u}{1 - \bar{u}z}\right| = \left|\frac{z - u}{\bar{u}(u - z)}\right| = 1.$$

Therefore, if we let $k \to \infty$ in (10), we get

$$\lim_{k \to \infty} R_k(z) = \frac{|1 - \bar{u}z|^2}{2 - 2\mathrm{Re}(\bar{u}z)} = \frac{|1 - \bar{u}z|^2}{1 - |z|^2 + |1 - \bar{u}z|^2} \stackrel{\text{def}}{=} R(z); \tag{12}$$

and, if we let $k \to \infty$ in (11), we get

$$\lim_{k \to \infty} C_k(z) = \frac{(1 - |z|^2)u}{2 - 2\mathrm{Re}(\bar{u}z)} = \frac{(1 - |z|^2)u}{1 - |z|^2 + |1 - \bar{u}z|^2} \stackrel{\text{def}}{=} C(z). \tag{13}$$

If we let $k \to \infty$ in (8), we get from (12) and (13)

$$|f^{[n]}(z) - C(z)| \leqslant R(z) \qquad \forall z \in D, n = 0,1,2,\ldots, \tag{14}$$

since evidently, for each fixed $z \in D$ and $n \in \mathbb{N}$ the nth iterate of f_k converges to the nth iterate of f. Notice that

$$|C(z)| = \frac{1 - |z|^2}{1 - |z|^2 + |1 - \bar{u}z|^2} < 1,$$

$$R(z) > 0 \quad \text{and} \quad |C(z)| + R(z) = 1.$$

Hence the closed disc of center $C(z)$ and radius $R(z)$ lies in $\bar{D}$ and is tangent to ∂D at the point

$$C(z) + R(z)\frac{C(z)}{|C(z)|} = C(z) + R(z)u = u.$$

According to (14) each point $f^{[n]}(z)$ lies in this disc; that is, for each $z \in D$ the closed disc D_z which contains z on its boundary and is internally tangent to ∂D at u also contains all the iterates $f^{[n]}(z)$. As any closed disc which contains z and is internally tangent to ∂D at u, contains D_z, the theorem is proved.

4. Iterates of Automorphisms. We can now completely describe the convergence behavior of the sequence $\{f^{[n]}\}$ when f is a conformal automorphism of D. According to 2.1, f has the form uT_a for some $a \in D, u \in \partial D$. We assume $f \neq I$. Then, as noted in 2.5, there are three cases to consider.

CASE I. *f has two (distinct) fixed points on ∂D. In this case the sequence $\{f^{[n]}\}$ converges locally uniformly in D to one of these points, namely, to the one farthest from $1/\overline{f^{-1}(0)}$.*

Proof. Let the fixed points be z_1 and z_2, so labeled that the number $\lambda = (1 - \bar{a}z_2)/(1 - \bar{a}z_1)$ satisfies $|\lambda| < 1$. (See 2.6.) Then, according to 2.4,

$$\frac{f^{[n]}(z) - z_1}{f^{[n]}(z) - z_2} = \lambda^n \cdot \frac{z - z_1}{z - z_2} \qquad \forall z \in D, n \in \mathbb{N}.$$

It follows that

$$|f^{[n]}(z) - z_1| \leqslant |\lambda|^n \cdot |f^{[n]}(z) - z_2| \left|\frac{z - z_1}{z - z_2}\right| \leqslant \frac{4}{1 - |z|}|\lambda|^n.$$

The right side converges to 0 with $1/n$, uniformly in any compact set of $z \in D$.

CASE II. *f has exactly one fixed point on ∂D. In this case the sequence $\{f^{[n]}\}$ converges locally uniformly in D to this fixed point.*

Proof. Letting z_0 denote the fixed point, it follows from 2.7 that

$$\frac{z_0}{f^{[n]}(z) - z_0} = n\left(\frac{u-1}{u+1}\right) + \frac{z_0}{z - z_0},$$

$$\frac{1}{|f^{[n]}(z) - z_0|} \geqslant n\left|\frac{u-1}{u+1}\right| - \frac{1}{|z - z_0|} \qquad \forall z \in D, \quad n \in \mathbb{N}.$$

Since $f \neq I$, it follows that $u \neq 1$. If K is a compact subset of D, then $m_K = \inf\{|z - z_0| : z \in K\}$ is positive and for all $n > \dfrac{2}{m_K}\left|\dfrac{u+1}{u-1}\right|$ our inequality yields

$$\frac{1}{|f^{[n]}(z) - z_0|} \geqslant n\left|\frac{u-1}{u+1}\right| - \frac{1}{m_K} > \frac{n}{2}\left|\frac{u-1}{u+1}\right|,$$

$$|f^{[n]}(z) - z_0| \leqslant \frac{2}{n}\left|\frac{u+1}{u-1}\right| \qquad \forall z \in K.$$

CASE III. *f has a unique fixed point z_0 in D. In this case either f is periodic in the sense that $f^{[n]} = I$ for some n, or the orbit $\{f^{[n]} : n \in \mathbb{N}\}$ is dense in the compact group $\mathcal{G}$ of all conformal automorphisms of D which fix z_0.*

Proof. Let $T = T_{z_0}$ and form $F = T \circ f \circ T^{-1}$. This is a conformal automorphism of D which fixes 0 and so, by 2.1, it has the form wI for some $w \in \partial D$. If w is an nth root of unity, then $f^{[n]} = T^{-1} \circ F^{[n]} \circ T = T^{-1} \circ (w^n I) \circ T = I$. If w is not a root of unity, then $\{w^n : n \in \mathbb{N}\}$ is dense in ∂D. The group $\mathcal{H} = \{e^{i\theta} I : 0 \leqslant \theta \leqslant 2\pi\}$ is compact and the map $e^{i\theta} I \to T^{-1} \circ (e^{i\theta} I) \circ T$ is continuous and sends $\mathcal{H}$ onto $\mathcal{G}$. The dense subset $\{w^n I : n \in \mathbb{N}\}$ of $\mathcal{H}$ maps to the orbit $\{f^{[n]} : n \in \mathbb{N}\}$ of f, which is consequently dense in $\mathcal{G}$.

REMARKS. An alternative way of treating the iteration of fixed-point-free conformal automorphisms of D is to go over to the open upper half-plane U, which is conformally equivalent to D. It is not hard to show that every fixed-point-free conformal automorphism f of U determines a conformal automorphism T such that $F = T^{-1} \circ f \circ T$ has either the form $F(z) = az$ for all $z \in U$ or the form $F(z) = z + b$ for all $z \in U$, where a is positive and b is real. The behavior of $\{F^{[n]}\}$ is obvious and this can be carried back to the original function on D by pre- and post-composing with the appropriate conformal map of D onto U and its inverse.

5. Iterates of Nonautomorphisms. Surprisingly, if $f : D \to D$ is holomorphic but not a conformal automorphism, its iterates can only have constant limits. Theorem 3.1 will be used to show that only one constant is possible, and then an appeal to Corollary 1.7 will show that the whole sequence $\{f^{[n]}\}$ converges. We begin with:

5.1. THEOREM. *Let $f : D \to D$ be holomorphic. Suppose that there exist $n_1 < n_2 < \cdots$ such that $f^{[n_j]} \to I$ in D. Then f is a conformal automorphism of D.*

Proof. Use Montel's Theorem to find $j_1 < j_2 < \cdots$ such that $\{f^{[n_{j_k} - 1]}\}$ converges, say to g. Then $f^{[n_{j_k}]} = f^{[n_{j_k} - 1]} \circ f \to g \circ f$, so $I = g \circ f$. In particular, f is nonconstant, so $f(D)$ is open (Open Map Theorem). Moreover, $(f \circ g) \circ f = f \circ (g \circ f) = f$, so $f \circ g = I$ in the set $f(D)$. By the Uniqueness Theorem, from $f \circ g = I$ in $f(D) \subset D$, follows $f \circ g = I$ in D.

5.2. COROLLARY. *Let $f : D \to D$ be holomorphic but not a conformal automorphism of D. Then every subsequential limit of $\{f^{[n]}\}$ is constant.*

Proof. Suppose that contrariwise there exist $n_1 < n_2 < \cdots$ such that $f^{[n_j]} \to g$ with g

nonconstant. Then g is holomorphic (by Weierstrass's Theorem) and so $g(D)$ is open (by the Open Map Theorem). Evidently $g(D) \subset \bar{D}$. Hence, in fact, $g(D) \subset D$, the interior of $\bar{D}$. Set $m_j = n_{j+1} - n_j$ and cite Montel's Theorem to come up with a convergent subsequence $\{f^{[m_{j_k}]}\}$. Let h denote its limit, which is holomorphic by (1.5.) For each $z \in D$ the convergent sequence $\{f^{[n_j]}(z)\}$ together with its limit $g(z)$ is a compact subset of D and so $\{f^{[m_{j_k}]}\}$ converges to h *uniformly* on this set. It follows easily that $f^{[m_{j_k}]}(f^{[n_{j_k}]}(z)) \to h(g(z))$. But $f^{[m_j]}(f^{[n_j]}(z)) = f^{[m_j+n_j]}(z) = f^{[n_{j+1}]}(z) \to g(z)$. Thus $h = I$ in the nonvoid open subset $g(D)$ of D. It follows from the Uniqueness Theorem that $h = I$ throughout D and then the last theorem implies that f is a conformal automorphism of D.

The climax of our mini-theory of iteration is at hand:

5.3. THEOREM. *Let $f: D \to D$ be holomorphic and have a fixed point z_0. Suppose that f is not a conformal automorphism of D. Then $\{f^{[n]}\}$ converges to z_0. In particular, the fixed point is unique.*

Proof. If g and h are any two subsequential limits of $\{f^{[n]}\}$, then $g(z_0) = z_0 = h(z_0)$. Since g and h are each constant by 5.2, it follows that $g = h$. But then by the Corollary to Montel's Theorem the whole sequence $\{f^{[n]}\}$ converges to z_0.

5.4. THEOREM (**[24]** and **[4]**). *Let $f: D \to D$ be holomorphic and fixed-point free. Then $\{f^{[n]}\}$ converges to a unimodular constant.*

Proof. The case where f is a conformal automorphism of D is covered by Case I and Case II above, so we shall suppose that f is not a conformal automorphism of D. Let g be any subsequential limit of $\{f^{[n]}\}$. By 5.2, g is constant, say $g(z) = \alpha$ for all z. If $\alpha \in D$ and $f^{[n_j]} \to g$, then on the one hand $f(f^{[n_j]}(z)) \to f(g(z)) = f(\alpha)$ by continuity of f at α. On the other hand, $f(f^{[n_j]}(z)) = f^{[n_j]}(f(z)) \to g(f(z)) = \alpha$, since $\{f^{[n_j]}\}$ converges to g at *each* point of $\bar{D}$. It follows that $f(\alpha) = \alpha$, against one of the hypotheses on f. Hence $\alpha \notin D$. Evidently $\alpha \in \bar{D}$, so $|\alpha| = 1$. But then it is obvious from 3.1 that α must coincide with the u of that theorem. Thus u is the only possible subsequential limit of $\{f^{[n]}\}$, so by the Corollary to Montel's Theorem the whole sequence $\{f^{[n]}\}$ converges to u—*mirabile dictu.*

5.5. COROLLARY. *If $f: D \to D$ is holomorphic and fixes two distinct points of D, then $f = I$.*

Proof. If f is a conformal automorphism of D, this follows from 2.1 and 2.5. Otherwise it follows from 5.3.

5.6 EXERCISE (**[18]**). *Show that i^i converges. That is, the sequence $\{i_n\}$, defined inductively by $i_1 = i, i_{n+1} = i^{i_n} = e^{i\pi i_n/2}$, converges.*

Hints (**[13]**). The region $\Omega = \{z \in \mathbb{C} : \operatorname{Re} z > 0, \operatorname{Im} z > 0, |z| < 1\}$ is conformal to D. Show that the entire function $f(z) = e^{\pi i z/2}$ maps Ω into a proper subset of itself and that the point $a = f^{[4]}(1)$ lies in Ω. Note that $i_n = f^{[n]}(1) = f^{[n-4]}(a)$.

5.7. THEOREM. *If the holomorphic map $f: D \to D$ leaves some nonempty compact subset K of D invariant, i.e., $f(K) \subset K$, then f has a fixed point.*

Proof. $f(K) \subset K \subset D$ causes all the iterates $f^{[n]}(z)$ to remain well inside D, if $z \in K$. In particular $\overline{\lim}|f^{[n]}(z)| \leq \sup|K| < 1$. Therefore, by 5.4, f is not fixed-point free.

6. Common Fixed Points. We start with a recent result of Theodore Mitchell **[14]** which complements Case III above.

6.1. THEOREM. *Let $\mathcal{G}$ be a group (composition) of holomorphic self-maps of D. Suppose $\mathcal{G}$ is compact (with respect to locally uniform convergence). Then either $\mathcal{G}$ consists of a single constant function or $\mathcal{G}$ consists of all powers of a single conformal automorphism of D which has finite order or $\mathcal{G}$ consists of all the conformal automorphisms of D which fix some one point of D. In any case, $\mathcal{G}$ is commutative and has a unique fixed point in D.*

Proof. Let E denote the identity element of $\mathcal{G}$. If E is a constant function, say $E(z) = z_0 \in D$ for every $z \in D$, then $f(z) = (E \circ f)(z) = E(f(z)) = z_0$ for all $f \in \mathcal{G}$. Thus $\mathcal{G} = \{E\} = \{z_0\}$.

If E is not constant, then $E(D)$ is an open subset of D (Open Map Theorem) and from $E(w) = (E \circ E)(w) = E(E(w))$ we infer that $E = I$ in the open set $E(D)$. By the Uniqueness Theorem then $E = I$ throughout D. It follows that the functions in $\mathcal{G}$ are all one-to-one and surjective, i.e., are conformal automorphisms of D. For any $I \neq f \in \mathcal{G}$, there are then only three possibilities:

(i) f has a unique fixed point in D.
(ii) f has a unique fixed point in $\bar{D}$, which lies on ∂D.
(iii) f has two distinct fixed points in $\bar{D}$, and they each lie on ∂D.

In the latter two cases the sequence $\{f^{[n]}\}$ in $\mathcal{G}$ converges in D to one of the fixed points (see §4), i.e., to a constant, hence to a function not in $\mathcal{G}$. This contradicts the compactness of $\mathcal{G}$. Thus (i) prevails for every $f \in \mathcal{G}$ different from I. If there are any such functions, let f_0 be one and let a_0 be its unique fixed point in D and set

$$T(z) = \frac{z + a_0}{1 + \bar{a}_0 z}.$$

Then set $F_0 = T^{-1} \circ f_0 \circ T$ and $\mathcal{G}_0 = T^{-1} \circ \mathcal{G} \circ T$, another compact group of conformal automorphisms of D. (See 2.1.) Now $F_0 \in \mathcal{G}_0$ and F_0 fixes 0. Therefore (2.1)

$$F_0(z) = e^{i\theta} z \qquad \text{for some } 0 < \theta < 2\pi. \tag{1}$$

Since also $F_0^{-1}(z) = e^{-i\theta} z = e^{(2\pi - \theta)i} z$ is in $\mathcal{G}_0$, we can replace θ by $2\pi - \theta$ and thereby assume that

$$0 < \theta \leqslant \pi. \tag{2}$$

Let $f \in \mathcal{G}_0$ maximize $|f(0)|$. Since $f \to f(0)$ is continuous and $\mathcal{G}_0$ is compact, such f exist. Write (after 2.1):

$$f(z) = u \frac{z + a}{1 + \bar{a} z}, \text{ where } |u| = 1 > |a|.$$

Write $u = e^{i\psi}$. Since $0 < \theta/\pi \leqslant 1$, there is an integer n such that

$$n \cdot \frac{\theta}{\pi} \in \left[-\frac{1}{2} - \frac{\psi}{\pi}, \frac{1}{2} - \frac{\psi}{\pi} \right].$$

Thus

$$\frac{n\theta + \psi}{\pi} \in \left[-\frac{1}{2}, \frac{1}{2} \right],$$

so

$$\operatorname{Re}(e^{in\theta} u) = \operatorname{Re}(e^{i(n\theta + \psi)}) \geqslant 0. \tag{3}$$

Set $w = e^{in\theta} u$ and note that

$$\operatorname{Re} w \geqslant 0 > -\frac{1 + |a|^2}{2} = -\frac{1}{2} \cdot \frac{1 - |a|^4}{1 - |a|^2},$$

$$2(1 - |a|^2) \operatorname{Re} w > -(1 - |a|^4),$$

$$2\operatorname{Re}\left[(|a|^2 - 1) w \right] < 1 - |a|^4$$

$$|a|^4 + 2\operatorname{Re}(w|a|^2) < 2\operatorname{Re} w + 1$$

$$|w|^2 |a|^4 + 2\operatorname{Re}(w|a|^2) + 1 < |w|^2 + 2\operatorname{Re} w + 1,$$

since $|w| = 1$,

$$|1 + w|a|^2|^2 < |1 + w|^2. \tag{4}$$

Now consider $f_1 = F_0^{[n]} \circ f = e^{in\theta} f$, an element of $\mathcal{G}_0$. We have

$$f_1^{[2]}(0) = f_1(f_1(0)) = f_1(wa) = w \cdot \frac{wa + a}{1 + \bar{a}wa} = wa \cdot \frac{1 + w}{1 + w|a|^2}.$$

Therefore

$$|f_1^{[2]}(0)| = |a|\left|\frac{1 + w}{1 + w|a|^2}\right| = |f(0)|\left|\frac{1 + w}{1 + w|a|^2}\right|. \tag{5}$$

From (4) and (5) we infer that $|f(0)| = 0$, for otherwise the maximality of $|f(0)|$ would be contradicted. It follows that $g(0) = 0$ for *every* $g \in \mathcal{G}_0$; i.e., 0 is the unique fixed point of each $g \in \mathcal{G}_0$.

In view of 2.1, $\mathcal{G}_0$ is a compact subgroup of $\{wI : w \in \partial D\}$; hence it is, in particular, commutative. But it is elementary that a subgroup of ∂D is either finite and cyclic or dense. Therefore either $\mathcal{G}_0$ is the totality of conformal automorphisms which fix 0 or $\mathcal{G}_0 = \{w^k I : k \in \mathbb{N}\}$ for some nth root w of 1. It follows that $\mathcal{G} = T \circ \mathcal{G}_0 \circ T^{-1}$ is commutative and is either the totality of conformal automorphisms which fix $T(0)$ or the cyclic subgroup $\{T \circ (w^k I) \circ T^{-1} = (T \circ (wT^{-1}))^{[k]} : k = 1,2,\ldots,n\}$ of the latter.

If inverses are lacking (i.e., the group hypothesis is dropped), but commutativity is available by fiat, then a common fixed point can be found by other arguments:

6.2. THEOREM ([**19**]). *Let $\mathcal{F}$ be a family of continuous self-maps of $\bar{D}$ which are holomorphic in D and commute with each other under composition. Then the functions in $\mathcal{F}$ have a common fixed point.*

Proof. We may evidently discard I from $\mathcal{F}$; i.e., suppose $I \notin \mathcal{F}$. Moreover, if $\mathcal{F}$ contains a constant function, then this constant is fixed by every element of $\mathcal{F}$, due to the commutativity. Hence we may also assume that each f in $\mathcal{F}$ is nonconstant. By the Open Map Theorem then $f(D)$ is an open subset of $\bar{D}$; so $f(D) \subset D$, for each $f \in \mathcal{F}$.

If some $g \in \mathcal{F}$ has a fixed point z_0 in D, then by 5.5 this fixed point is unique. Since $g(f(z_0)) = f(g(z_0)) = f(z_0)$ for each $f \in \mathcal{F}$, uniqueness of z_0 implies that the g-fixed point $f(z_0)$ coincides with z_0. Thus z_0 is a fixed point of every $f \in \mathcal{F}$.

If, on the other hand, no g in $\mathcal{F}$ fixes any point of D, we look at any particular $g \in \mathcal{F}$. According to 5.4 there is a unimodular complex number u such that $g^{[n]}(z) \to u$ for every $z \in D$. But for any $f \in \mathcal{F}$ we also have $z = f(0) \in f(D) \subset D$; so $f(u) = \lim_n f(g^{[n]}(0)) = \lim_n g^{[n]}(f(0)) = u$. It follows that u is a fixed point of every $f \in \mathcal{F}$.

7. Historical Remarks. In [**23**] Wolff proved 5.4 under the additional assumption that f have a continuous extension to $\bar{D}$. A few weeks later Wolff [**24**] and Denjoy [**4**] independently succeeded in removing this hypothesis. A proof similar to Wolff's was later given by Valiron [**21**], [**22**]. Here is an outline of it: The set of subsequential limits of $\{f^{[n]}\}$ is a set of unimodular constants; this we showed in the course of proof of 5.4. It is easy to prove that this set is connected, hence is an arc of ∂D. In case f is continuous on $\bar{D}$ the method of 5.4 also shows that f fixes each point in this arc. A very elementary boundary uniqueness theorem asserts that $f = I$ if this arc has positive length. Since f has no fixed point in D, the arc is a single point and is a fixed point of f. In the absence of continuity on ∂D, this method would seem doomed. However, there is a rather deep theorem of Fatou [**6**] which asserts that for almost every (in the sense of Lebesgue measure) $\theta \in [0,2\pi]$, $f(z)$ has a limit as z approaches $e^{i\theta}$ inside any triangle in D with vertex at $e^{i\theta}$ (so-called *nontangential* approach). Calling this limit suggestively $f(e^{i\theta})$, Fatou also showed that, if for two such functions f and g, $f(e^{i\theta}) = g(e^{i\theta})$ almost everywhere in a nondegenerate arc of ∂D, then $f = g$. The proof sketched above can now be pushed through. With more effort it also leads to an extension of 6.2 (due to Behan [**2**]) in which the hypothesis of continuity on $\bar{D}$ is dropped. To formulate it, note that the unimodular constant $u = u_f$ of 5.4 is evidently unique. Wolff showed that $f(z) \to u_f$ as $z \to u_f$ nontangentially, so that there is a natural extension of f to a function on $D \cup \{u_f\}$ which fixes u_f and is "almost" continuous. Behan's extension of 6.2 reads:

if g is another such function, which is not a Case I automorphism, and $g \circ f = f \circ g$, then $u_g = u_f$. The original proof of 6.2 uses fewer results from function theory but instead relies in an interesting way on elementary facts about compact topological semigroups. Similar techniques lead to higher dimensional extensions of 6.2. See Eustice [**5**] and Suffridge [**20**].

The proof of 5.4 based on 3.1 was discovered only a little later by Wolff [**25**]. It was inspired by Denjoy [**4**] and Julia [**10**, pp. 72–77], [**11**]. A somewhat similar treatment was subsequently offered by Craig and Macintyre [**3**] and a half-plane transcription of Wolff's proof appears in the last section of Montel [**15**]; but this second proof of Wolff's has nevertheless been overlooked by many later writers. The case of rational f (mapping D into D), which subsumes the results of §4, was announced by Fatou in [**7**] and treated by him in Chapter 3 of [**8**].

For extensions to multiply-connected regions see Heins [**9**].

The opening phase of the proof of 3.1 illustrates a more general result of Ritt [**17**]: If Ω is a region and f a holomorphic function in Ω such that $f(\Omega)$ lies in a compact subset of Ω, then $\{f^{[n]}\}$ converges to a constant. This too has higher dimensional analogs.

For more on the theme of 5.6, iterating complex exponents, see Shell [**18**] and the papers in his bibliography.

Mitchell's version of 6.1 asserts only the existence of the common fixed point and the proof is different from that in the text. I thank Professor Mitchell for making available to me a preprint of his work, from which I evolved the treatment above. Theorem 5.7 is an easy corollary of a deeper result about left reversible semigroups of self-maps of D which he proves by an elegant reduction to 6.1 via elementary topological semigroup arguments.

References

1. J. J. Andrews and R. C. Lacher, Stacked exponents, Aequationes Math., 16 (1977) 137–147. **MR** 57 #6331.

2. D. F. Behan, Commuting analytic functions without fixed points, Proc. Amer. Math. Soc., 37 (1973) 114–120. **MR** 46 #7492.

3. C. Craig, Jr., and A. J. Macintyre, Inequalities for functions regular and bounded in a circle, Pac. Jour. of Math., 20 (1967) 449–454. **MR** 34 #7806.

4. A. Denjoy, Sur l'itération des fonctions analytiques, C. R. Acad. Sci Paris, 182 (1926) 255–257. **Jbuch** 52, p. 309.

5. D. J. Eustice, Holomorphic idempotents and common fixed points on the 2-disk, Mich. Math. Jour., 19 (1972) 347–352. **MR** 47 #2040.

6. P. Fatou, Séries trigonométriques et séries de Taylor, Acta Math., 30 (1906) 335–400. **Jbuch** 37, p. 283.

7. ______, Sur les substitutions rationnelles, C. R. Acad. Sci. Paris, 164 (1917) 806–808. **Jbuch** 46, p. 519.

8. ______, Sur les équations fonctionnelles, Bull. Soc. Math. France, 47 (1919) 161–271 and 48 (1920) 33–94 and 208–314. **Jbuch** 47, p. 921.

9. M. H. Heins, On the iteration of functions which are analytic and single-valued, in a given multiply-connected region, Amer. Jour. of Math., 63 (1941) 461–480. **MR** 2, p. 275.

10. G. Julia, Mémoire sur l'itération des fonctions rationnelles, Jour. de Math. Pures et Appliq., (8) 1 (1918) 47–245. **Jbuch** 46, p. 520.

11. ______, Extension nouvelle d'un lemme de Schwarz, Acta Math., 42 (1920) 349–355. **Jbuch** 47, p. 272.

12. R. A. Knoebel, Exponentials reiterated, this MONTHLY, 88 (1981) 235–252.

13. A. J. Macintyre, Convergence of i^{i} , Proc. Amer. Math. Soc., 17 (1966) 67. **MR** 32 #5855.

14. T. Mitchell, Invariant means and analytic actions, Pac. Jour. of Math., 85 (1979) 145–153.

15. P. Montel, Sur les fonctions méromorphes limites de fractions rationnelles à termes entrelacés, Ann. Sci. École Norm. Sup., (3) 50 (1933) 171–196. **Jbuch** 59, p. 331.

16. G. Pólya and G. Szegő, Problems and Theorems in Analysis I (translated from German by D. Aeppli), Springer-Verlag, New York, Heidelberg, Berlin, 1972. **MR** 49 #8782.

17. J. F. Ritt, On the conformal mapping of a region into a part of itself, Annals of Math., (2) 22 (1920–21) 157–160. **Jbuch** 48, p. 363.

18. D. J. Shell, On convergence of infinite exponentials, Proc. Amer. Math. Soc., 13 (1962) 678–681. **MR** 25 #5307.

19. A. L. Shields, On fixed points of commuting analytic functions, Proc. Amer. Math. Soc., 15 (1964) 703–706. **MR** 29 #2790.

20. T. J. Suffridge, Common fixed points of commuting holomorphic maps of the hyperball, Mich. Math. Jour., 21 (1974) 309–314. **MR** 51 #3903.

21. G. Valiron, Sur l'itération des fonctions holomorphes dans un demi-plan, Bull. Sci. Math., (2) 55 (1931) 105–128. **Jbuch** 57, p. 381.

22. ______, Fonctions analytiques, Presses Universitaires de France, Paris, 1954. **MR** 15, p. 861.

23. J. Wolff, Sur l'itération des fonctions holomorphes dans une région, et dont les valeurs appartiennent à cette région, C. R. Acad. Sci. Paris, 182 (1926) 42–43. **Jbuch** 52, p. 309.

24. ______, Sur l'itération des fonctions bornées, C. R. Acad. Sci. Paris, 182 (1926) 200–201. **Jbuch** 52, p. 309.

25. ______, Sur une généralisation d'un théorème de Schwarz, C. R. Acad. Sci. Paris, 182 (1926) 918–920 and 183 (1926) 500–502. **Jbuch** 52, p. 309.

26. ______, Über die Iteration derjenigen in einem Gebiete regulären Funktionen, deren Werte dem Gebiete angehören, Math. Zeit., 26 (1927) 125–127. **Jbuch** 53, p. 304.

RIGID MOTIONS OF CONICS: AN INTRODUCTION TO INVARIANT THEORY

FRANK D. GROSSHANS

Department of Mathematical Sciences, West Chester State College, West Chester, PA 19380

Students of analytic geometry know that the conic defined by the equation

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0 \qquad (*)$$

is obtained from a "standard conic" via a rigid motion. In this paper, we consider the following question: Can the equation of this standard conic be found in terms of the coefficients a, b, c, d, e, and f?

The answer, first discovered about one hundred years ago, is given by invariant theory. We shall give a modern exposition using matrix techniques to explain and extend this old result. The Euclidean group is defined in §1 along with a suitable action on symmetric matrices. Three polynomials, denoted by τ, δ, and Δ, are shown to be invariant. The orbits of this action are considered in §2 where canonical forms are found. Conics are then reconsidered in §3. The exposition in §4 is entirely devoted to a proof that τ, δ, and Δ generate all the invariants of the group action defined earlier. Finally, the matrix techniques used in this paper are powerful and generalize easily. Some of these extensions are noted in §5.

1. A Group Action. (1.1) *The Euclidean group.* A 2×2 matrix N with entries in $\mathbb{R}$ is called *special orthogonal* if

$${}^tNN = 1 \quad \text{and} \quad \det N = 1.$$

It can be shown that such a matrix corresponds to a rotation through some angle θ and can be written

$$N = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The *Euclidean group*, which we shall denote by G, consists of all 3×3 real (partitioned) matrices of the form

$$\left(\begin{array}{cc|c} \multicolumn{2}{c|}{N} & B \\ \hline 0 & 0 & 1 \end{array}\right) \qquad (1)$$

where N is special orthogonal and B is a 2×1 column matrix. (To obtain the inverse of (1), we

Frank David Grosshans received his Ph.D. from the University of Chicago in 1967 under I. Satake. He has taught at DePaul University, the University of Pennsylvania, and Johns Hopkins University. He is now a Professor of Mathematical Sciences at West Chester State College in Pennsylvania. His research interests are in invariant theory. He dedicates this paper to his colleague Sebastian Koh.—*Editors*

just replace N by ${}^tN$ and B by $-{}^tNB$.) We note that each element in G has determinant 1.

(1.2) *Orbits of G on V.* For a moment, let V be an abstract set. We recall that an action of G on V is given by a mapping

$$G \times V \to V \text{ denoted by } (g,v) \to g\cdot v$$

such that (i) $(gh)\cdot v = g\cdot(h\cdot v)$ and (ii) $1\cdot v = v$ for all v in V and g, h in G. The *orbit* of an element v in V is

$$G\cdot v = \{g\cdot v : g \in G\}.$$

Now let V be the vector space consisting of all 3×3 real symmetric matrices. For $g \in G$ and $Q \in V$, we define

$$g\cdot Q = {}^tg^{-1}Qg^{-1}. \tag{2}$$

It is easy to verify that (2) gives an action of G on V.

Next, we wish to study the orbit under G of an element Q in V. To do this, let us denote a typical element Q in V by

$$Q = \begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix} = \left(\begin{array}{c|c} A(Q) & D(Q) \\ \hline {}^tD(Q) & f(Q) \end{array}\right) \tag{3}$$

where

$$A(Q) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad D(Q) = \begin{pmatrix} d \\ e \end{pmatrix}, \quad \text{and} \quad f(Q) = f.$$

To simplify matters let us denote the matrix (1) by g^{-1} and suppose that Q' is in the orbit of Q; that is, $Q' = {}^tg^{-1}Qg^{-1}$. Then, by a direct calculation, we find that

$$\begin{cases} A(Q') = {}^tNA(Q)N \\ D(Q') = {}^tNA(Q)B + {}^tND(Q) \\ f(Q') = {}^tBA(Q)B + 2{}^tD(Q)B + f(Q). \end{cases} \tag{4}$$

(1.3) *Invariant polynomials.* Let R be the polynomial ring in six indeterminates, $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6$, over the reals. Then R can be identified with the ring of polynomial functions on V as follows: for $P = P(Z_1,Z_2,Z_3,Z_4,Z_5,Z_6)$ in R and Q in V as in (3), put

$$P(Q) = P(a,b,c,d,e,f).$$

A polynomial P in R is called *invariant* with respect to the action of the Euclidean group if $P(g\cdot Q) = P(Q)$ for all g in G and Q in V. Let I be the set of all such invariant polynomials. Then I is an algebra over $\mathbb{R}$.

Next, we define three polynomials τ, δ, and Δ in R and show that each is invariant. Let

$$\tau(Q) = \operatorname{trace} A(Q), \quad \delta(Q) = \det A(Q), \quad \Delta(Q) = \det Q. \tag{5}$$

We see that that τ (respectively, δ) is invariant by taking the trace (respectively, determinant) of the equation $A(Q') = {}^tNA(Q)N$. The invariance of Δ follows by taking the determinant of both sides of $Q' = {}^tg^{-1}Qg^{-1}$.

Our goal in §4 is to show that any invariant polynomial in R can be written uniquely as a polynomial in τ, δ, and Δ. That is to say, we shall prove the following:

THEOREM. *If $\mathbb{R}[\tau,\delta,\Delta]$ denotes the subalgebra of R generated by τ, δ, and Δ, then $I = \mathbb{R}[\tau,\delta,\Delta]$. Furthermore, τ, δ, and Δ are algebraically independent over $\mathbb{R}$.*

2. Canonical Forms. Let Q be as in (3) and let $p(x)$ be the characteristic polynomial of $A(Q)$, i.e.,

$$p(x) = x^2 - \tau(Q)x + \delta(Q).$$

Let λ and μ be the characteristic values of $A(Q)$. There is a special orthogonal matrix N so that ${}^tNA(Q)N$ is a diagonal matrix with diagonal entries λ and μ. (Indeed, choose θ so that $\tan 2\theta = 2b/(a-c)$.) Let g^{-1} have the form (1) with this N and $B=0$. According to (4), we then find in the orbit of Q the matrix

$$Q' = {}^tg^{-1}Qg^{-1} = \begin{pmatrix} \lambda & 0 & d' \\ 0 & \mu & e' \\ d' & e' & f \end{pmatrix}. \tag{6}$$

Furthermore, since $D(Q') = {}^tND(Q)$ we have

$$(d')^2 + (e')^2 = {}^tD(Q')D(Q') = {}^tD(Q)D(Q) = d^2 + e^2. \tag{7}$$

We may also assume that $e' \geqslant 0$ (otherwise, apply g^{-1} of the form (1) with N a rotation through π and $B=0$).

Case 1. If $\delta(Q) \neq 0$, then the orbit of Q contains the matrix

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \Delta(Q)/\delta(Q) \end{pmatrix}.$$

Such a diagonal matrix in the orbit of Q is uniquely determined up to an interchange of λ and μ.

Proof. Since $\delta(Q) = \delta(Q') = \lambda\mu \neq 0$, there exist s and t so that

$$\lambda s + d' = 0 \quad \text{and} \quad \mu t + e' = 0.$$

Let $Q'' = g\cdot Q'$ where g^{-1} has the form (1) with $N = 1$ and ${}^tB = (s,t)$. Then Q'' is diagonal with diagonal entries λ, μ, and f'', say. But

$$\Delta(Q) = \Delta(Q'') = \lambda\mu f'' = \delta(Q)f''$$

so that Q'' has the desired form. The uniqueness of the form follows from (4) since λ and μ must be the characteristic values of $A(Q)$. These characteristic values may be interchanged by taking N to be a rotation through $\pi/2$ and $B=0$.

Case 2. If $\delta(Q) = 0$, $\tau(Q) \neq 0$, $\Delta(Q) \neq 0$, then the orbit of Q contains the matrix

$$\begin{pmatrix} \tau(Q) & 0 & 0 \\ 0 & 0 & e'' \\ 0 & e'' & 0 \end{pmatrix}$$

where e'' is the positive square root of $-\Delta(Q)/\tau(Q)$.

Proof. Since $\delta(Q) = 0$, we may assume in (6) that $\lambda = \tau(Q)$ and $\mu = 0$. We note that $\Delta(Q) = \Delta(Q') = -\tau(Q')(e')^2$ so that $e' \neq 0$. To get the desired form, let $Q'' = g\cdot Q'$ where g^{-1} has the form (1) with $N = 1$, ${}^tB = (s,t)$ and

$$0 = \lambda s + d' \quad \text{and} \quad 0 = \lambda s^2 + 2(d's + e't) + f.$$

Case 3. If $\delta(Q) = 0$, $\tau(Q) \neq 0$, $\Delta(Q) = 0$, then the orbit of Q contains the matrix

$$\begin{pmatrix} \tau(Q) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f' \end{pmatrix}$$

where $f' = f - (d^2 + e^2)/\tau(Q)$.

Proof. Since $\delta(Q) = 0$, we may assume in (6) that $\lambda = \tau(Q)$ and $\mu = 0$. We also have $e' = 0$ since $\Delta(Q) = \Delta(Q') = 0$. To obtain the desired form, let $Q'' = g\cdot Q'$ where g^{-1} has the form (1) with $N = 1$ and ${}^tB = (s,0)$ with $\lambda s + d' = 0$. (Then use (4) and (7).)

Case 4. If $\delta(Q) = \tau(Q)\Delta = (Q) = 0$, then the orbit of Q contains either the matrix

$$\begin{pmatrix} 0 & 0 & d' \\ 0 & 0 & 0 \\ d' & 0 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f \end{pmatrix}$$

where $(d')^2 = d^2 + e^2$.

Sketch of proof. Since $\delta(Q) = \tau(Q) = 0$, we have $A = 0$. There is an N so that ${}^tND = {}^t(d'0)$. If $d' \neq 0$, we can get the matrix on the left.

3. Rigid Motions of Conics. In this section, it will be convenient to identify the (ordinary) plane with the plane $z = 1$ in $\mathbb{R}^3$ so that

$$p = \begin{pmatrix} x \\ y \end{pmatrix} \leftrightarrow P = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}.$$

Then, if $(*) Q(p) = Q(x,y) = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f$ and Q is as in (3) we have

$$Q(p) = Q(x,y) = {}^tPQP.$$

Let us denote the matrix (1) by g. The Euclidean group sends the plane $z = 1$ to itself so that g gives rise to the rigid motion F_g where $F_g(p) = gP = Np + B$.

Now let $Q' = g\cdot Q$ and let C and C' be the conics associated to Q and Q'; i.e., $C = \{p : Q(p) = 0\}$ and $C' = \{p : Q'(p) = 0\}$. The equation

$${}^tPQP = {}^tP{}^tgQ'gP = {}^t(gP)Q'(gP)$$

shows that p lies on C if and only if $gP = F_g(p)$ is on C'. Thus F_g sends C to C'. The description of canonical forms in §2 gives rise immediately to the following answers to the question posed at the start of this paper.

Case 1. $\delta(Q) \neq 0$. Let λ and μ be the roots of the equation $x^2 - \tau(Q)x + \delta(Q) = 0$. The standard conic from which equation (*) is obtained by a rigid motion corresponds to

$$\lambda x^2 + \mu y^2 + \Delta(Q)/\delta(Q) = 0.$$

Case 2. $\delta(Q) = 0$, $\tau(Q) \neq 0$, $\Delta(Q) \neq 0$. Let e' be the positive square root of $-\Delta(Q)/\tau(Q)$. Then the standard conic is

$$\tau(Q)x^2 + 2e'y = 0.$$

Case 3. $\delta(Q) = 0$, $\tau(Q) \neq 0$, $\Delta(Q) = 0$. The standard conic is

$$\tau(Q)x^2 + f - (d^2 + e^2)/\tau(Q) = 0.$$

Case 4. $\delta(Q) = \tau(Q) = \Delta(Q) = 0$ but $d^2 + e^2 \neq 0$. The standard conic is $(d^2 + e^2)^{1/2}x = 0$.

4. The Ring of Invariants. (4.1) *A lemma.* In this section, we shall prove the theorem stated in (1.3). To make the exposition as self-contained as possible, we first prove a lemma (which is the case $n = 2$ of the Fundamental Theorem of Symmetric Functions).

LEMMA. *Let x and y be indeterminates and let $S = xy$ and $T = x + y$. Let $F(x,y)$ be a polynomial over $\mathbb{R}$ so that $F(x,y) = F(y,x)$. Then there is a polynomial $G(x,y)$ over $\mathbb{R}$ so that $F(x,y) = G(S,T)$.*

Proof. First, let $F(x,y) = x^k + y^k$. Then the lemma follows by induction on k from the equation

$$x^k + y^k = T(x^{k-1} + y^{k-1}) - S(x^{k-2} + y^{k-2}).$$

Next, we may suppose that $F(x,y)$ is homogeneous, say,

$$F(x,y) = a_0x^n + a_1x^{n-1}y + \cdots + a_{n-1}xy^{n-1} + a_ny^n.$$

Since $F(x,y) = F(y,x)$ we have $a_0 = a_n, a_1 = a_{n-1}, \ldots$; so

$$F(x,y) = a_0(x^n + y^n) + a_1S(x^{n-2} + y^{n-2}) + \cdots$$

and we use the fact just proved.

(4.2) *A homomorphism.* Let us define a subset X of V by

$$X = \left\{ \begin{pmatrix} a & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & f \end{pmatrix} = [a,c,f] : ac \neq 0 \right\}.$$

Also, let U be the subset of V defined by $U = \{Q \in V : \delta(Q) \neq 0\}$. Any element in U can be transformed by the Euclidean group to an element in X according to Case 1, §2. As usual, we shall identify polynomials in R with the corresponding polynomial functions on V and polynomials in $\mathbb{R}[Z_1,Z_3,Z_6]$ with polynomial functions on X.

We now define a mapping $H: I \to \mathbb{R}[Z_1,Z_3,Z_6]$ by

$$P(Z_1,Z_2,Z_3,Z_4,Z_5,Z_6) \to P(Z_1,0,Z_3,0,0,Z_6).$$

Alternatively, if we think of P as a function on V, then

$$P \to P|X = \text{the restriction of } P \text{ to } X.$$

The mapping H is a homomorphism. Furthermore, H is one-to-one. For if $H(P) = 0$, then $P|X = 0$ and, since P is invariant, $P|U = 0$. Since U is dense in V, we have $P = 0$. Let us conclude this section by noting that

$$H(\tau) = Z_1 + Z_3, \quad H(\delta) = Z_1Z_3, \quad H(\Delta) = Z_1Z_3Z_6.$$

(4.3) *The basic equation.* Let P be polynomial in I. There are unique polynomials $g_0(Z_1,Z_3), \ldots, g_m(Z_1,Z_3)$ such that

$$H(P) = g_0(Z_1,Z_3)Z_6^m + \cdots + g_m(Z_1,Z_3). \tag{8}$$

Now we saw in Case 1, §2, that there is an element in the Euclidean group which sends the element $[a,c,f]$ in X to the element $[c,a,f]$ in X. Since P is invariant, we conclude that

$$P(a,0,c,0,0,f) = P(c,0,a,0,0,f)$$

for all a,c,f such that $ac \neq 0$. Hence, $g_i(Z_1,Z_3) = g_i(Z_3,Z_1)$ for each $i = 0,\ldots,m$. According to the lemma in (4.1), each $g_i(Z_1,Z_3)$ is itself a polynomial in $S = Z_1Z_3$ and $T = Z_1 + Z_3$, say, $g_i(Z_1,Z_3) = h_i(S,T)$. Rewriting (8) we obtain

$$H(P) = h_0(S,T)Z_6^m + \cdots + h_m(S,T). \tag{9}$$

From (9), we will derive the basic equation, namely,

$$\delta^m P = h_0(\delta,\tau)\Delta^m + h_1(\delta,\tau)\delta\Delta^{m-1} + \cdots + h_m(\delta,\tau)\delta^m. \tag{10}$$

To do this, we first observe that both sides of (10) are invariant polynomials. We also recall that H is a one-to-one homomorphism such that $H(\tau) = T$ and $H(\delta) = S$. Now applying H to each side of (10) gives $(Z_1Z_3)^m H(P)$.

(4.4) *The last step.* The polynomials τ, δ, and Δ are algebraically independent since their images under H are. Therefore, we may rewrite the right-hand side of (10) and get

$$\delta^m P = \delta^e\big(j_0(\tau,\Delta)\delta^k + \cdots + j_k(\tau,\Delta)\big) \tag{11}$$

where $j_k(\tau,\Delta) \neq 0$. If $e \geqslant m$, we may cancel δ^m from both sides of (11) to see that P is a polynomial in τ, δ, and Δ.

So let us suppose that $e < m$ in (11) and seek a contradiction. Let $i = m \quad e$. Then

$$\delta^i P = j_0(\tau,\Delta)\delta^k + \cdots + j_k(\tau,\Delta) \tag{12}$$

where $i > 0$. Since $j_k(\tau,\Delta) \neq 0$, there are real numbers α and β so that $j_k(\alpha,\beta) \neq 0$ and $\alpha\beta < 0$. Let $\gamma = (-\beta/\alpha)^{1/2}$ and put

$$Q = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & \gamma \\ 0 & \gamma & 0 \end{pmatrix}.$$

Evaluating both sides of (12) at Q gives

$$0 = j_k(\alpha,\beta) \neq 0,$$

which is a contradiction.

5. Generalizations. An $n \times n$ matrix N with entries in $\mathbb{R}$ is called *special orthogonal* if

$${}^tNN = 1 \quad \text{and} \quad \det N = 1.$$

The (extended) *Euclidean group*, denoted by G, consists of all $(n+1) \times (n+1)$ real (partitioned) matrices of the form (1) where N is an $n \times n$ special orthogonal matrix and B is an $n \times 1$ column matrix.

Now let V be the vector space consisting of all $(n+1) \times (n+1)$ real symmetric matrices. We denote a typical element Q in V by

$$Q = \left(\begin{array}{c|c} A(Q) & D(Q) \\ \hline {}^tD(Q) & f(Q) \end{array}\right)$$

where $A(Q)$ is an $n \times n$ (symmetric) matrix, $D(Q)$ is an $n \times 1$ column matrix, and $f(Q)$ is 1×1. The group G acts on V via $g \cdot Q = {}^tg^{-1}Qg^{-1}$ and equations (4) hold. A matrix Q as above gives rise to a quadratic form on $\mathbb{R}^n$. Indeed, let x be the $n \times 1$ column vector with ${}^tx = (x_1,\ldots,x_n)$ and put

$$Q(x_1,\ldots,x_n) = {}^txA(Q)x + 2{}^tD(Q)x + f(Q).$$

Most of what was done in earlier sections extends (without much trouble) to the setting just given. In fact, we have the following theorems:

(1) *Let us write the characteristic polynomial of $A(Q)$ as*

$$\det(x1 - A(Q)) = x^n - \delta_1(Q)x^{n-1} + \delta_2(Q)x^{n-2} + \cdots + (-1)^n\delta_n(Q)$$

where each δ_i is a polynomial in the coefficients of $A(Q)$. Then, the polynomials $\delta_1,\ldots,\delta_n$ are invariant under the action of the Euclidean group. Furthermore, $\Delta(Q) = \det Q$ is also an invariant polynomial.

(2) *Under the action of G, the form $Q(x_1,\ldots,x_n)$ associated to Q can be brought to one of the following canonical forms*:

$$\lambda_1 x_1^2 + \cdots + \lambda_r x_r^2 + f' \qquad (r \leqslant n)$$

or

$$\lambda_1 x_1^2 + \cdots + \lambda_r x_r^2 + 2d'x_{r+1} \qquad (d' > 0, r < n)$$

where $\lambda_1,\ldots,\lambda_r$ are the nonzero characteristic values of $A(Q)$. This form is uniquely determined up to a permutation of $\lambda_1,\ldots,\lambda_r$.

(3) *Case* 1 *in* §2 *still holds. That is, if* $\delta_n(Q) \neq 0$, *then* $Q(x_1,\ldots,x_n)$ *can be brought to the form*

$$\lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 + \Delta(Q)/\delta_n(Q)$$

where $\lambda_1,\ldots,\lambda_n$ *are the characteristic values of* $A(Q)$.

(4) *The algebra I of invariant polynomials of G acting on V is generated by* $\delta_1,\ldots,\delta_n$, *and* Δ. *Furthermore, these polynomials are algebraically independent.*

The theorems just stated answer three basic questions in the study of invariant theory.

I. *Finite generation*: Is the algebra of invariant polynomials finitely generated (over the base field)? This is, essentially, Hilbert's fourteenth problem; most of the known affirmative cases rely on nonconstructive arguments. In our example, (4) answers this question by explicitly displaying the invariants.

II. *Algebraic relations*: What is the structure of the ring of invariants; what are the algebraic relations among the generators? In our example, (4) answers the question again by stating that I is a polynomial ring.

III. *Quotient spaces*: Can orbits be separated by invariant functions, i.e., if v_1 is not in the orbit of v_2, is there an invariant polynomial P so that $P(v_1) \neq P(v_2)$? In our example, (3) provides an answer: as long as we restrict our attention to the open set $U = \{Q : \delta_n(Q) \neq 0\}$, we may separate orbits by invariant polynomials.

Annotated Bibliography

1. A. D. Aleksandrov, et al., Mathematics: Its Content, Methods, and Meaning, 2nd. ed., MIT Press, Cambridge, Mass., 1963. The chapter on analytic geometry is excellent and states the Theorem of (1.3).

2. J. Dieudonné and J. B. Carrell, Invariant Theory, Old and New, Academic Press, New York, 1971. The focus is on Questions I and III. Some examples are given and the symbolic method is explained.

3. J. Fogarty, Invariant Theory, W. A. Benjamin, New York, 1969. This is an introductory text that explains the modern approach to Questions I and III.

4. J. Humphreys, Hilbert's fourteenth problem, this MONTHLY, 85 (1978) 341–353. This is a survey of progress on Question I.

5. F. Klein, Elementary Geometry from an Advanced Standpoint: Geometry, Dover, New York, 1945. The middle part of this book is on geometry and invariant theory.

6. M. Nagata, Lectures on the Fourteenth Problem, Tata Institute of Fundamental Research, Bombay, 1965. This gives a thorough examination of Question I, including the counterexample found in 1958 by Nagata.

7. G. Salmon, A Treatise on Conic Sections, 6th ed., Chelsea, New York, 1954. This is the classic work on conics. The invariance of τ and δ is proved but the structure of I is not given. Rather, Salmon gives geometric interpretations of various invariants (especially in Chapter 18).

8. T. A. Springer, Invariant Theory, Springer-Verlag, New York, 1977. This gives a modern treatment of Question I along with the examples of SL_2 and finite-reflection groups. A review of this book by R. Gardner is given in: Bulletin A.M.S., 2 (1980) 246–256. The reviewer's bibliography indicates applications of invariant theory to many diverse areas of mathematics.

9. H. Weyl, The Classical Groups, Princeton University Press, Princeton, N.J., 1946. This modern classic looks at Questions I and II in the concrete setting of the classical groups.

WOMEN AND MATHEMATICS: FACT AND FICTION

EDITH H. LUCHINS

Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12181

The thesis that cultural and social factors are relevant to the underrepresentation of women in mathematics seems to be gaining acceptance [6], [7], [17], [25]. Such factors are regarded as playing significant roles in discouraging women from entering mathematics and also in discouraging those who do from devoting themselves single-mindedly to it. Yet interviews with mathematicians,

For a biographical note, see this MONTHLY, 86 (1979) 161.—*Editors*

psychologists, and educators [**16**] and various current writings (e.g., [**11**], [**19**], [**23**], [**32**]) suggest that considerable attention is also being paid to noncultural and nonsocial factors, to factors which are regarded as innate or as relatively resistant to change, in attempts to account for the scarcity of women mathematicians or for sex differences in mathematics.

Often cited as contributing to sex differences in mathematics are differences in various aptitudes or abilities. In particular, it is widely reported that males excel in spatial ability and females in verbal ability [**5**], [**18**], [**19**], [**28**]. It is of interest that these were the second and third most frequently cited sex differences, as revealed in a survey of over 100 general psychology texts, covering the century from 1875 to 1975; the survey was conducted by my husband and his students when he was Visiting Professor at the University of Guelph. See Table 1 for the most frequently cited sex differences. The only sex difference that was cited more frequently than those for spatial and verbal abilities was males' greater physical strength, which was cited in every text. Males' superior spatial ability was cited by 91 percent of the texts, and females' superior verbal aptitude in 83 percent of them. Males were described as better with numbers or with computations in 79 percent of the books. Females were reported to talk earlier in 76 percent of them. Generations of students, teachers, and counselors have had such stereotypes about sex differences drummed into them—even when scientific evidence did not exist! An example is the belief that women are more emotional than men, which was cited in 100 percent of the surveyed texts up to 1960, but in only 22 percent since then.

TABLE 1. Frequency of Sex Differences Cited in Over 100 General Psychology Texts (1875–1975)

	Cited in % of texts
Males—physically stronger	100.0
Males—better spatial ability	91.3
Females—better verbal ability	82.7
Males—better mechanical aptitude	81.7
Females—more neurotic	81.7
Males—better with numbers	78.8
Females—better mental imagery	75.9
Females—talk earlier	75.9
Females—better memory	75.0
Males—more interested in aesthetics	69.2

This table is adapted from study by Abraham S. Luchins and his students at the University of Guelph, Guelph, Ontario, of over 100 general psychology texts in the university library which were published in the century from 1875 to 1975.

In their comprehensive survey of sex differences, Maccoby and Jacklin [**18**] conclude that differences in spatial ability and in quantitative ability are not clear-cut in the early years. When sex differences in these abilities are found in the age range of 9 to 13 years, they tend to favor boys. Beginning with adolescence, boys tend to move ahead of girls in spatial and in quantitative abilities. However, some recent studies are not finding clear-cut sex differences in spatial abilities (e.g., [**1a**], [**6a**]).

Sex differences in verbal development are depicted by Maccoby and Jacklin [**18**] as having three phases. One phase occurs early, before 3 years of age, when girls seem to have an advantage. At about age 3, the boys tend to catch up and, in most population groups, the two sexes have similar verbal performances until about 11 years of age. (However, where there are differences from ages 3 to 11, they favor girls; these differences tend to be found in underprivileged populations.) A new phase of differentiation generally occurs at adolescence, with the girls

showing verbal superiority. Thus, beginning in adolescence, if not earlier, females tend to have an advantage for verbal ability and males for spatial ability.

What are the determinants of these sex differences? Innate factors are often appealed to, especially for spatial ability. Thus a 1977 text on the psychology of women notes that, of the three areas of fairly well established sex differences, "spatial ability appears to be most likely to have an innate [and inheritable] component" [**32**, p. 137].

One theory is that spatial ability is a recessive trait carried on the X-chromosome [**4**], [**5**]. Let X_0 denote the absence and X_1 the presence of spatial ability. Since X_1 is recessive, spatial ability will be manifested if X_1 is present and if X_0 is not present. According to this theory, the only way for a female, who normally has two X-chromosomes, to manifest spatial ability is if she received X_1 from both parents. If she received it from only one parent, then the dominant factor on the other X-chromosome would suppress it. On the other hand, if a boy received X_1 from his mother, then he would manifest the trait (whether or not his mother manifested it), since from his father he could receive only the Y-chromosome, so that there would be no dominant factor on the X-chromosome to suppress spatial ability. Thus there is a higher probability that males would show spatial ability; that is, about half of the males and a quarter of the females should manifest it [**4**].

One difficulty with this hypothesis is that it assumes that spatial ability is an all-or-nothing affair—either one has it or not—whereas it seems to lie on a continuum. Moreover, the hypothesis accepts the assumptions that one is born that way, that the factor is innate, and that the distribution is the same in all cultures, thereby ignoring some evidence that training and culture can affect spatial ability [**3**], [**26**]. Furthermore, strong evidence for the X-chromosome hypothesis has not been found in recent investigations. A 1979 survey article by McGee of human spatial ability concludes that "the X-linked recessive gene hypothesis that has served as a tentative explanation for sex differences in spatial abilities and for the mode of genetic transmission is not supported strongly in recent studies" [**19**, p. 909]. A 1980 critical review [**4a**] rejects the hypothesis as unfounded.

Other hypotheses focus on hormonal differences, e.g., that the lack of the sex hormone estrogen accounts for males' better spatial abilities. Research findings in this area are highly ambiguous (see the surveys in [**18**], [**19**]). For instance, Petersen [**22**] investigated the relation of the physical manifestations of sex hormones to cognitive functioning and found that the relatively less-masculine (androgynous) physical characteristics were *positively* related to spatial ability and *negatively* related to fluent (verbal) productions. This is opposite to what would be expected from the hormonal hypothesis. There is further evidence that highly masculinized males tend to have lower spatial scores [**19**]. It would be of interest to see what happens to spatial ability with changes, natural or induced, in estrogen levels, e.g., of post-menstrual women.

The most prevalent hypotheses to account for differences in spatial ability are those which appeal to sex differences in brain functioning. In a popular 1979 book on the brain [**23**], which has received considerable attention in the press [**24**], the neurologist Richard Restak states, "Recent psychobiological research indicates that many of the differences in brain function between the sexes are innate, biologically determined, and relatively resistant to change through the influence of culture" [**23**, p. 197]. Differences in cognitive functions and, in particular, sex differences in spatial (and verbal) ability, are attributed to differences in the specialization, maturation, or dominance of one or the other cortical hemispheres, referred to as cortical laterality [**10**], [**11**], [**19**], [**29**]. There are several related hypotheses.

In most right-handed people, the left cortical hemisphere controls the right hand and gains information from the right visual field, while the reverse is true for the right hemisphere. The left cortical hemisphere specializes in language or verbal processes while the right is concerned with nonverbal information and spatial tasks. There are also believed to be characteristic modes of operation. The left hemisphere favors sequential analysis, breaking a task into component parts or details and dealing with them separately, while the right favors pattern analysis, a holistic, Gestalt approach.

Females are believed to develop specialization or lateralization earlier or more strongly than males. Earlier dominance of their left hemisphere would account for their speaking earlier, for their alleged earlier sensitivity to sound, superior verbal aptitude, and superiority in sequential tasks. Concomitantly, the right hemisphere seems to become less dominant and tends to be inhibited by the left. This would help to explain females' alleged poorer spatial ability. On the grounds that we tend to favor the better hemisphere, it might also account for female mathematicians choosing algebra as a speciality more often than a speciality in which spatial perception is more central (at least this was the case for the women mathematicians in our NSF study [**16**]).

Many reports in the literature conclude that right hemispheric dominance or specialization is greater in males than in females (see the reviews in [**7**] and [**29**]). This would account for males' alleged superiority in spatial tasks which call for a holistic approach and also for their poorer verbal performances. EEG measurements recently made of electrical events in the brain suggest that boys tend to use their right hemisphere for spatial tasks and girls their left hemisphere for both spatial and verbal tasks. This results in a kind of interference effect for girls, a deleterious effect of the use of words on solutions of spatial problems [**10**], [**11**], [**23**].

The situation is by no means clear-cut with regard to the role played in mathematical abilities by the different hemispheres and by spatial and verbal factors. We all know mathematicians who have poor spatial ability and yet do very good mathematics. Different branches of mathematics, e.g., algebra and geometry, seem to call for different degrees of spatial ability. In general, mathematics involves both spatial and verbal processes as well as both analytic and holistic processes. Hence mathematics should utilize both cerebral hemispheres. Restak [**23**], who attributes sex differences in spatial ability to differences in brain laterality, and who assigns spatial processes to the right hemisphere, nonetheless claims that "the *left* hemisphere controls written and spoken language *and* mathematical abilities" (p. 167, italics added) and describes mathematics as a *left*-hemisphere subject (p. 184). Also, the left hemisphere is generally assumed to control logical, analytical, and rational processes which surely are involved in mathematics. If females have better developed left cortical hemispheres, then why are they not better in mathematics?

Two psychologists, Martha Mednick and Nancy Felipe Russo [**20**], contend that Restak offers a misleading review of the literature. Their main arguments are as follows: Studies with infants have not conclusively demonstrated that either sex is more visual or more auditory than the other [**18**]. Differences which are found later on are not universal or unmodifiable. Female verbal superiority was not found in Germany or in Israel. Also, sex differences in spatial ability were not found for Canadian Eskimos for whom it has considerable survival value [**3**]. Granted that boys and girls in our society show differences in spatial ability beginning at age 9 or later, what about the training that boys have received in tasks involving such abilities, e.g., through their toys and games? The critique concludes that perhaps the most consequential mistake Restak makes is in his confusion between groups and individuals, particularly as it affects his prescriptions for educational reform.

The present author also believes that there is cause for concern in the apparent uncritical acceptance by some mathematicians and educators of the assumed sex differences in brain functioning—and the educational practices they would base on them. At the joint meeting held at the University of Minnesota, Duluth, in August 1979, one mathematical educator proposed that the class be divided by sex, with different methods of teaching used for females and for males, because of differences in their brain functioning. Such proposals seemingly overlook that the alleged sex differences in cortical functioning are not firmly established and that their relationships to spatial, verbal, and mathematical abilities, or to differences in these abilities, are not clear. Moreover, neither the hypotheses about the differences nor the research to date contain intrinsic educational implications (cf. [**28**]). Even if it were the case that, in a large population, females would tend to be better in verbal ability and males in spatial ability, these group trends need not hold for a given female or a given male. If we are going to go out of our way to differentiate educational curricula and methods on the basis of verbal or spatial ability, then why should sex be

the criterion? If the only information available about the students was their sex, then we would have to be limited to it. But surely more fine-grained information can be obtained directly about spatial, verbal, and mathematical abilities and then used as the basis for possible educational changes. See [**6a**] and [**17**] for discussions of educational implications of sex differences in mathematics.

It has been suggested that females have been deprived of learning experiences which facilitate spatial ability [**26**], [**30**]. Perhaps more stress on spatial factors and less stress on verbal presentation of mathematics would, possibly in different ways, help both sexes.

Pilot studies are needed to explore changes that can be made in methods of teaching, and in the content and organization of the curriculum, to take into account possible differences (and not necessarily sex differences) in mathematical aptitudes and achievements. Studies are also needed to explore the effects of differences in attitudes and interests in mathematics and in what has been called styles of thinking. Our investigation [**13**], using two-dimensional geometric problems that require spatial, Gestalt visualization, suggests that sex differences may be less pronounced than other differences. In one experiment the boys had 26 percent more solutions than the girls, but the female mathematics majors had 51 percent more solutions than the female nonmathematics majors. Our experiments on three-dimensional figures [**14**] suggest that sex differences may be less important than differences in methods of presentation. When the task given to high school students was to recognize a geometric solid from its description, it was found that contra-structural descriptions (those which did not follow the natural division of the structure of the solid) yielded statistically significant sex differences that favored the boys. Pro-structural descriptions resulted in significantly more solutions for both boys and girls than did contra-structural descriptions, with the change more marked for the girls. Pro-structural descriptions virtually eliminated sex differences. See Table 2 for results obtained with descriptions of a tetrahedron. The study hints that it may be more fruitful to find structurally appropriate methods of teaching mathematics than to argue over possible sex differences in mathematics. More generally, methods of teaching mathematics may make for greater variations in learning than do sex differences.

TABLE 2. Recognition of a Solid (Tetrahedron) by Male and Female High School Students

Subjects	Percent Recognition from Descriptions Contra-Structural	Pro-Structural	Difference	t-value
	%	%	%	
Males (N = 79)	68	94	26	4.91*
Females (N = 63)	35	87	52	7.07*

Description	Percent Recognition from Subjects Males (N = 79)	Females (N = 63)	Difference	t-value
	%	%	%	
Contra-Structural	68	35	33	4.14*
Pro-Structural	94	87	7	1.40†

*t-value denotes statistical significance at better than the 0.0006 level.
†t-value denotes that the difference is not statistically significant at acceptable levels.

While spatial ability of students has been studied for over half a century [**18**], [**19**], little is known about the spatial ability of mathematicians. Let us turn to some preliminary observations that we made in the summer of 1979 at an applied mathematics conference, where females were conspicuous by their scarcity. This exploratory study used a questionnaire that asked questions on spatial ability and also on handedness; the two have been related in the literature (e.g., [**11**], [**18**], [**27**]). The mathematicians were asked whether they used the right or the left hand in a variety of situations (cf. [**1**], [**21**]). Of 30 males at the conference, over one-quarter reported that they were

either left-handed or ambidextrous (which is higher than the approximately 10 to 14 percent estimated for the general public). When asked to rate their ability to visualize spatial structures, 93 percent of the applied mathematicians rated themselves as very good or good. Ratings of very good were assigned to their spatial abilities by 50 percent of the non-right-handed mathematicians but by only 18 percent of the right-handed ones. Two-thirds of the group said that spatial ability often helped them in mathematics and almost one-third said that it helped sometimes. We have been interested in how the answers to such questions would compare for other applied mathematicians, for pure mathematicians, for male and female mathematicians, for mathematicians in various specialties, and for other creative individuals, e.g., artists and musicians. We are undertaking research in attempts to answer some of these questions. The results to date point to some relationships between left-handedness and such factors as sex, order of birth, and season of birth (cf. [**2**], [**8**], [**9**]).

Half of the mathematicians at the applied mathematics conference were born in the four months from November through February. It is also in these months that most prenatal stress is believed to occur. Moreover, schizophrenics without a family history of schizophrenia are born more frequently in these months [**31**]. Furthermore, there is a relatively high incidence of left-handedness among schizophrenics [**12**]. A recent study using CAT scans (computerized X-ray topography of the brain) showed some tendency for reversals of laterality among certain schizophrenics, e.g., the right-handed ones were less likely to show the brain asymmetry found in normal right-handers [**15**]. One hypothesis is that mixed laterality (lack of clear-cut lateralization) predisposes to left-handedness and also to psychological stresses such as learning disorders, behavior disorders, and schizophrenia. Possibly it also predisposes to mathematical creativity and other forms of creativity. Perhaps mixed laterality may prove to be the link between madness and genius which was once thought to be an old wives' tale.

But what are we as mathematicians and educators supposed to do about all these hypotheses? Fascinating as they are, we do not know if they are fanciful or fruitful. It will be of interest to know what further research reveals about them and whether the findings help to separate fact from fiction about mathematicians in general and female mathematicians in particular.

This paper is based on an MAA invited address delivered at the Joint Meeting held at the University of Minnesota, Duluth, Minnesota, on August 22, 1979.

References

1. Marian Annett, A classification of hand preference by association analysis, British J. of Psychology, 61 (1970) 303-321.

1a. Jane Armstrong, Achievement and Participation of Women in Mathematics: An Overview, Education Commission of the States, Denver, Colo., 1980.

2. P. Baken, Handedness and birth order, Nature, 229 (1971) 195.

3. J. W. Berry, Temne and Eskimo perceptual skills, International J. of Psychology, 1 (1966) 207-229.

4. D. R. Bock and D. Kolakowski, Further evidence of sex-linked major-gene influence on human spatial visualizing ability. Amer. J. Human Genetics, 25 (1973) 1-14.

4a. D. B. Boles, X-linkage of spatial abilities: A critical review, Child Development, 51 (1980) 625-635.

5. A. W. H. Buffery and J. A. Gray, Sex differences in the development of spatial and linguistic skills, in C. Ounsted and D. C. Taylor, eds., Gender Differences: Their Ontogeny and Significance, Williams & Wilkins, Baltimore, 1972.

6. John Ernest et al., Mathematics and sex, this MONTHLY, 83 (1976) 595-614.

6a. Elizabeth Fennema, The sex factor: Real or not in mathematics education, in Elizabeth Fennema, ed., Implications of Mathematics Education Research on the Curriculum of the 80's, Association for Supervision and Curriculum Development, Washington, D.C. (to appear).

7. Irene H. Frieze, Women and Sex Roles: A Social Psychological Perspective, W. W. Norton, 1978.

8. Alan Leviton and Ted Kilty, Birth order and left-handedness, Archives of Neurology, 33 (1976) 664.

9. ______, Seasonal variation in the birth of left-handed schoolgirls, Archives of Neurology, 36 (1979) 115-116.

10. Jerre Levy, Possible basis for the evolution of lateral specialization of the human brain, Nature, 22 (1969) 614–615.

11. ______, Cerebral lateralization and spatial ability, Behavior Genetics, 6 (1976) 171–188.

12. W. A. Lishman and E. R. L. Meekan, Hand preference patterns in psychiatric patients, British J. of Psychiatry, 129 (1976) 158–166.

13. Abraham S. Luchins and Edith H. Luchins, Geometric problem solving related to differences in sex and mathematical interests, J. of Genetic Psychology, 134 (1979) 255–269.

14. ______, Differences between pro-structural and contra-structural descriptions of geometric figures outweigh sex differences (in preparation).

15. Daniel J. Luchins, Daniel R. Weinberger, and Richard J. Wyatt, Schizophrenia: Evidence of a subgroup with reversed cerebral asymmetry, Archives of General Psychiatry, 36 (1979) 1309–1311.

16. Edith H. Luchins, Women in Mathematics: Problems of Orientation and Reorientation, Final Report, NSF Grant #GY 11316, 1976.

17. ______, Sex differences in mathematics: How *not* to deal with them, this MONTHLY, 86 (1979) 161–168.

18. Eleanor E. Maccoby and Carol N. Jacklin, The Psychology of Sex Differences, Stanford Univ. Press, Stanford, Calif., 1974.

19. Mark G. McGee, Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences, Psychological Bulletin, 80 (1979) 889–918.

20. Martha Mednick and Nancy Felipe Russo, Boys vs. girls: Is there really a difference? Washington Post; reprinted in the Times-Union, Albany (Sunday, August 5, 1979).

21. R. C. Oldfield, The assessment and analysis of handedness: The Edinburgh inventory, Neuropsychologia, 9 (1971) 97–113.

22. Anne C. Petersen, Physical androgyny and cognitive functioning in adolescence, Developmental Psychology, 12 (1976) 524–533.

23. Richard Restak, The Brain: The Last Frontier, Doubleday, New York, 1979.

24. ______, Like it or not, biology rules boy-girl roles, Washington Post, reprinted in the Times-Union, Albany (Sunday, July 1, 1979). The other difference between boys and girls, Reader's Digest, November 1979, p. 214.

25. Lucy Sells, Sex and discipline differences in doctoral attrition, Ph.D. thesis, University of California, Berkeley, 1975.

26. J. A. Sherman, Problem of sex differences in space perception and aspects of intellectual functioning, Psychological Review, 74 (1967) 290–299.

27. ______, Cognitive performance as a function of sex and handedness: An evaluation of the Levy hypothesis, Psychology of Women Quarterly (to appear).

28. I. M. Smith, Spatial Ability: Its Educational and Social Significance, University of London, 1964; Knapp, San Diego, Calif., 1964.

29. D. C. Taylor, Differential rates of cerebral maturation between sexes and between hemispheres, Lancet, 2 (1969) 140–142.

30. Sheila Tobias, Overcoming Math Anxiety, W. W. Norton, New York, 1978.

31. E. F. Torrey, B. B. Torrey and M. R. Patterson, Seasonality of schizophrenic birth in the U.S.A., Archives of General Psychiatry, 34 (1978) 1065–1070.

32. Juanita H. Williams, Psychology of Women: Behavior in a Biosocial Context, Norton, New York, 1977.

MISCELLANEA

56.

On account of the empirically verified fact that at this point in the contemporary time frame we no longer find ourselves in an inability situation with regard to the eventual realization of our desired goals, and in view of the effectively secured acquisition of the requisite capability...

—Anonymous Committee Report

Because now we can...

THE GEOMETRY OF ROLLING CURVES

JOHN BLOOM AND LEE WHITT
Department of Mathematics, Texas A & M University, College Station, TX 77843

Roll a closed convex curve along a line and follow the path of any chosen point on the curve. In the simplest case, the well-known cycloid is traced by a point on a rolling circle. In general, the set of (pointed) closed convex curves produces a wide variety of traced curves. Which curves are produced this way? Given a curve, can it be traced by rolling a (pointed closed convex) curve? If so, which one? In this paper, we give the necessary and sufficient conditions for traceability in terms of the normals to the curve and construct the curve to be rolled.

Suppose that the tracing point is allowed to be inside or outside of the rolling curve. Suppose further that the "line of roll" is replaced by a curve and that nonconvex curves are allowed to roll, i.e., by requiring that the point of contact move smoothly with no sliding (arclengths must agree) and the tangent lines agree at the contact point. In all cases, we solve the local inverse problem, as before, in terms of the normals to the curve.

The geometry of rolling curves has been studied extensively by mechanical engineers and others (see bibliography; Besant's book is the earliest systematic study) but their solution of the inverse problem is somewhat incomplete. We thank Dr. Rundell for suggesting this problem. All curves are plane curves. For simplicity, C^∞ differentiability is assumed unless explicitly stated otherwise. This topic may be suitable for an honors calculus class.

1. Necessary Condition. We begin with a simple case. Let C be a closed convex curve which can *roll* along a line L, i.e., the curvature of C is positive except possibly on a nowhere dense set and so there are no "straight sides." This condition guarantees that the point of contact s is well defined and behaves as the arclength parameter on both C and L (cf. §3). The tracing point P can be placed inside, on, or outside C. These are illustrated in Fig. 1 with a circle.

If P is regarded as the origin, then C can be described by polar coordinates (r,θ). If ψ is the angle between the tangent line and radial line of C, then P traces out the curve $\bar{C}$ (Fig. 2) given by

$$x = s - r\cos\psi \qquad y = r\sin\psi \tag{1}$$

We claim that the radial line is normal to $\bar{C}$. Using the well-known equation $\tan\psi = r\,d\theta/dr$, we calculate

$$\begin{aligned} ds &= \sqrt{dr^2 + r^2 d\theta^2} \\ &= \sqrt{1 + \left(\frac{r\,d\theta}{dr}\right)^2}\, dr \\ &= \sqrt{1 + \tan^2\psi}\, dr \\ &= \sec\psi\, dr, \end{aligned}$$

and so

$$\frac{dx}{d\psi} = \frac{ds}{d\psi} - \frac{dr}{d\psi}\cos\psi + r\sin\psi = \sin\psi\left(r + \tan\psi\frac{dr}{d\psi}\right)$$

and

$$\frac{dy}{d\psi} = \cos\psi\left(r + \tan\psi\frac{dr}{d\psi}\right).$$

John Bloom completed his doctoral work at Caltech under the tutelage of H. Kisilevsky in 1977 and has been teaching at Texas A & M ever since. His research interests include algebraic geometry, number theory, combinatorics, and numerical analysis.

Lee Whitt received his Ph.D. in 1975 from Yale under R. Szczarba. He spent two years at Cornell before moving to his present position at Texas A & M. His main research interest is differential geometry.—*Editors*

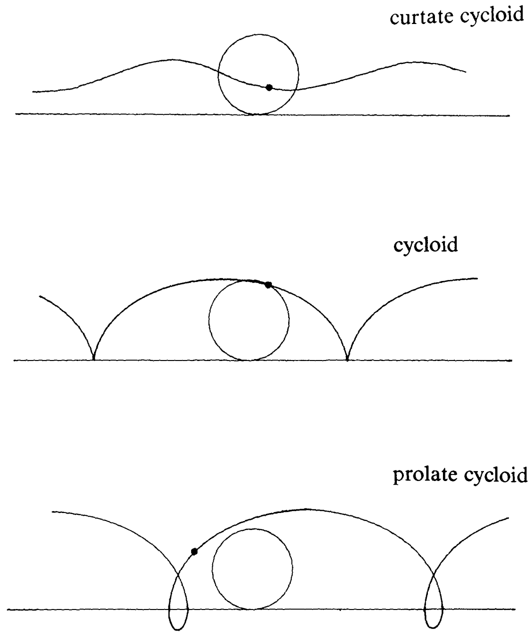


FIG. 1

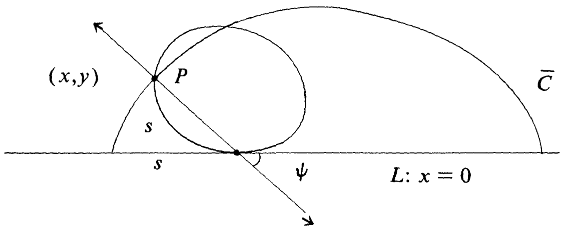


FIG. 2

Hence $dy/dx = \cot\psi$ and the radial line is normal to $\bar{C}$. In particular,

CONDITION 1. The normals to $\bar{C}$ intersect L in increasing order.

This condition is sufficient for the local construction of C (Lemma 1).

An important observation is that the angle sum $\theta + \psi$ is the angle between the tangent lines of C and a fixed line (the polar axis). Since $d(\theta + \psi)/ds$ is the curvature of C by definition, the curvature assumption on C is equivalent to $d(\theta + \psi)/ds \geqslant 0$, with equality only on a nowhere dense set.

CONDITION 2. The function $(1/y)(dx/ds)$ is positive except possibly on a nowhere dense subset of $\bar{C}$.

It suffices to show

$$\frac{1}{y}\frac{dx}{ds} = \frac{d\theta}{ds} + \frac{d\psi}{ds}. \tag{2}$$

By construction,

$$\frac{y}{s-x} = \tan\psi \quad \text{and} \quad \frac{dy}{dx} = \frac{s-x}{y}.$$

So,

$$\frac{d\psi}{dx} + \frac{d\theta}{dx} = \frac{d}{dx}\left(\tan^{-1}\frac{y}{s-x}\right) + \frac{\tan\psi}{r}\frac{dr}{dx}$$

$$= \frac{\frac{dy}{dx}(s-x) - y\left(\frac{ds}{dx} - 1\right)}{(s-x)^2 + y^2} + \frac{r\tan\psi}{r^2}\frac{dr}{dx}$$

$$= \frac{\frac{(s-x)^2}{y} - y\sec\psi\frac{dr}{dx} + y + r\tan\psi\frac{dr}{dx}}{r^2}.$$

Since $y\sec\psi = r\tan\psi$, we obtain

$$\frac{d\psi}{dx} + \frac{d\theta}{dx} = \frac{1}{y}\left[\frac{(s-x)^2 + y^2}{r^2}\right] = \frac{1}{y},$$

which is equivalent to equation (2).

Equation (2) has several amusing interpretations. From the curvature assumption on C, it follows that P moves forward above L and backward below L. This is equivalent to the popular brain-teaser: What part of a train is moving backward? The answer is: the part of the inner wheel flange that drops below the track. Also, if $\bar{C}$ crosses L, then it crosses orthogonally, i.e., $\theta + \psi$ is defined everywhere and so $dx|_{\bar{C}} = 0$ whenever $y = 0$.

If nonconvex curves are rolled, then the position of P (above or below L) and the curvature of C at the contact point determine the direction P travels. It follows that the nonconvexity of C can be an obstruction to the smoothness of $\bar{C}$. For example, if C is nonconvex and if $\bar{C}$ lies above L, then $(1/y)dx$ changes sign and P moves forward and backward above L. If $\bar{C}$ is smooth, then it has a vertical tangent line. But the normal lines to $\bar{C}$ must intersect L, and hence $\bar{C}$ cannot be smooth.

2. Sufficient Conditions. Let $\bar{C}$ be a plane curve which satisfies condition 1 and which is periodic with respect to a line L. In this way, the arclength parameter s on L also parametrizes $\bar{C}$. The differentiability assumption on $\bar{C}$ is that the length $r(s)$ of the normal vector from $\bar{C}$ to L is smooth and $(dr/rds)\tan\psi$ (or equivalently $(1/y)(dx/ds) - (d\psi/ds)$) has a smooth extension over all its singularities. These smoothness conditions do not imply that $\bar{C}$ is smooth.

LEMMA 1. *If L and $\bar{C}$ are as above, then there is a smooth closed, not necessarily convex or simple, curve C and a distinguished point P which traces any finite piece of $\bar{C}$.*

Proof. Let r be the length of the normal vector from $\bar{C}$ to L, and let θ satisfy the differential equation

$$\frac{d\theta}{ds} = \frac{dr}{rds}\tan\psi.$$

The curve given by $(r(s),\theta(s))$ is smooth by assumption. To see that this curve is C, it suffices to check that the arclength parameter of $(r(s),\theta(s))$ is s.

Since equations (1) still hold, it follows that

$$\frac{dy}{d\psi} = \cos\psi\left(r + \tan\psi\frac{dr}{d\psi}\right) \quad \text{and} \quad \frac{dy}{dx} = \cot\psi.$$

Hence

$$\frac{dx}{d\psi} = \sin\psi\left(r + \tan\frac{dr}{d\psi}\right)$$

and, on differentiating equations (1),

$$\begin{aligned}\frac{ds}{d\psi} &= \sec\psi\frac{dr}{d\psi}\\ &= \sqrt{1+\tan^2\psi}\,\frac{dr}{d\psi}\\ &= \sqrt{1+\left(\frac{rd\theta}{dr}\right)^2}\,\frac{dr}{d\psi}.\end{aligned}$$

Hence $ds = \sqrt{dr^2 + (rd\theta)^2}$ and s is the arclength of C. If our construction of C does not result in a smooth closed curve, then we easily complete C to obtain the desired curve (Figs. 4 and 5). Q.E.D.

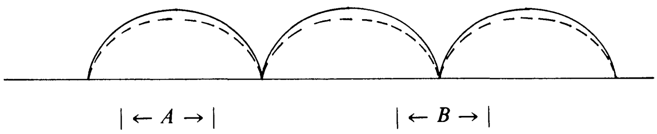


FIG. 3

The local construction of C contains several peculiarities. For example, if all of $\bar{C}$ is used, and if $\bar{C}$ is the graph of $y = \log x$, $x \geqslant 1$, then C is a spiral. Another example is best illustrated by a curve $\bar{C}$ (solid curve) lying slightly above a cycloid (dotted curve) (Fig. 3). Our main theorem requires a period of $\bar{C}$, and it is possible for different periods to produce different curves C. Over a period like A, our construction of C produces a corner at the origin (solid curve) which is then completed (dotted curve) to a smooth closed curve (Fig. 4). Over a period like B, our construction of C produces a nonclosed curve which is then completed (Fig. 5). This dependence on the period can be resolved by introducing the following integral condition (3) into our main theorem.

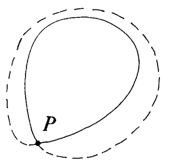


FIG. 4

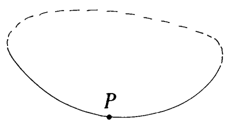


FIG. 5

THEOREM 1. *Let L and $\bar{C}$ be as in Lemma 1. Assume further that $\bar{C}$ satisfies Condition 2 and*

$$\int_{\bar{C}_1} \frac{1}{y}\,dx = 2\pi \tag{3}$$

where $\bar{C}_1$ is a period of $\bar{C}$ relative to L. Then there is a unique smooth closed convex curve C and a distinguished point P which traces all of $\bar{C}$.

Proof. At the endpoints of $\bar{C}_1$, the normals agree. Also, the change in ψ on $\bar{C}_1$, the integral condition (3), and equation (2) determine the change in θ. Namely, if ψ changes by 0, π, or 2π, then θ changes by 2π, π, or 0, respectively, where these correspond to, for example, the prolate, ordinary, and curtate cycloids in Fig. 1. If θ changes by 2π or 0, then applying Lemma 1 to $\bar{C}_1$ produces a closed curve C. If ψ changes by π then, by the periodicity of $\bar{C}$, r must be zero at the ends and also C is closed. The smoothness of C follows from the continuity (equivalently, closure) of C and the smoothness of the tangent vector $(dr/ds, d\theta/ds)$ to C. The convexity follows from Condition 2 and equation (2). Finally, C is unique up to a rotation of the (r,θ)-plane because r is

uniquely determined by $\bar{C}$ and θ is defined up to a constant, or equivalently, a rotation. Q.E.D.

3. Relaxation of Some Conditions. As mentioned earlier, a nonconvex curve C can roll under a suitable definition of "roll," and the traced curve $\bar{C}$ will not be smooth. The nonsmoothness of $\bar{C}$ can also be caused by flat sides on C. Such sides force a discontinuity in the normal vector field to $\bar{C}$. If a square is rolled along a line and the tracing point P is chosen to be a corner, then we obtain Fig. 6. Here we clearly see the effects of corners and flat sides. When C rolls at a corner, the normals meet the line in a stationary fashion. The flat sides cause the singularities. Notice also that $\bar{C}$ is not convex, even though C is. If C is rounded slightly to produce a smooth convex curve, the traced curve $\bar{C}$ will still be nonconvex, although it will now be smooth.

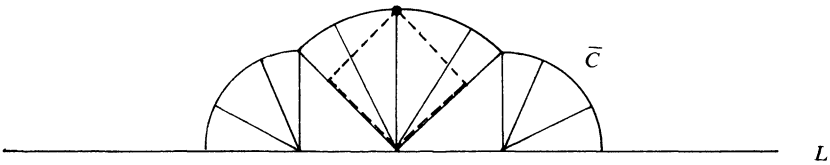


FIG. 6

4. Rolling Curves Along Curves. Let a curve C with tracing point P (not necessarily on C) be rolled along a curve $\underline{C}$ to produce a traced curve $\bar{C}$ (Fig. 7). We parametrize these curves by (r,θ), (z,u), and (x,y), respectively. Let α be the angle between the normal to $\underline{C}$ and the axis.

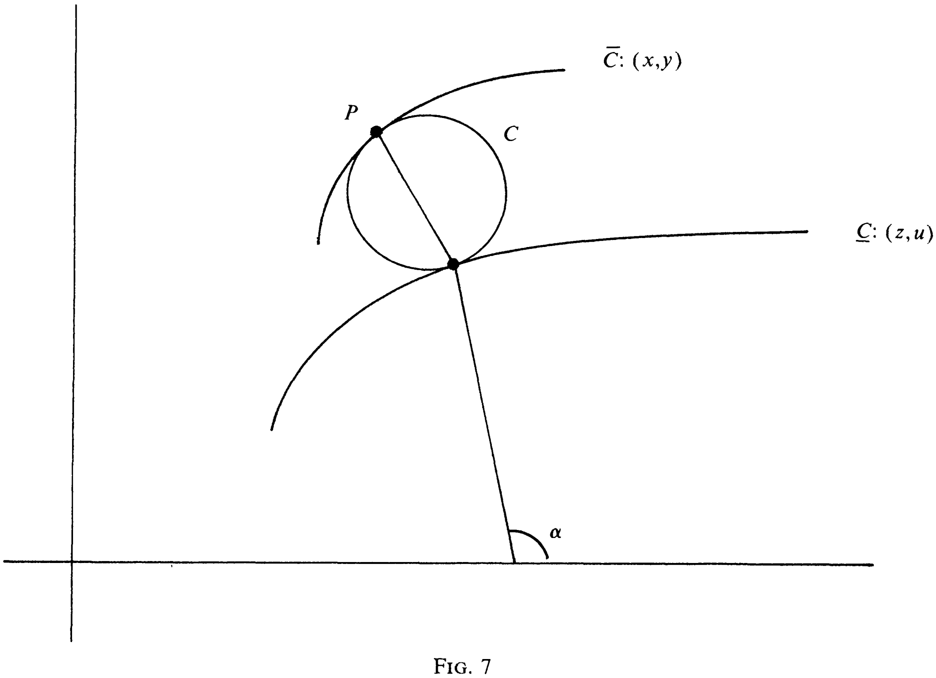


FIG. 7

We claim that, as before, the line between P and the contact point is perpendicular to $\bar{C}$. Consider

$$\begin{aligned} x &= r\cos(90° + \alpha - \psi) + z = -r\sin(\alpha - \psi) + z \\ y &= r\sin(90° + \alpha - \psi) + u = r\cos(\alpha - \psi) + u \end{aligned} \tag{4}$$

where ψ is the angle between the normal to $\bar{C}$ and the tangent to $\underline{C}$ at the contact point. Since the arclength parameter on $\underline{C}$ agrees with that on C, we obtain

$$\sqrt{dz^2 + du^2} = ds = \sec\psi\, dr.$$

From $du/dz = \tan(90° + \alpha)$ and $dz/du = -\tan\alpha$, it follows that

$$\sec\psi\, dr = -\sec\alpha\, du = \csc\alpha\, dz.$$

Differentiating equation (4),

$$\begin{aligned}\frac{dx}{d\alpha} &= -\frac{dr}{d\alpha}\sin(\alpha-\psi) - r\cos(\alpha-\psi)\left(1-\frac{d\psi}{d\alpha}\right) + \frac{dz}{d\alpha}\\ &= -\frac{dr}{d\alpha}\sin(\alpha-\psi) - r\cos(\alpha-\psi)\left(1-\frac{d\psi}{d\theta}\right) + \sin(\alpha-\psi+\psi)\sec\psi\frac{dr}{d\alpha}\\ &= \cos(\alpha-\psi)\left[-r\left(1-\frac{d\psi}{ds}\right) + \tan\psi\frac{dr}{d\alpha}\right].\end{aligned}$$

Similarly,

$$\frac{dy}{d\alpha} = \sin(\alpha-\psi)\left[-r\left(1-\frac{d\psi}{ds}\right) + \tan\psi\frac{dr}{d\alpha}\right],$$

and so

$$\frac{dy}{dx} = \tan(\alpha-\psi),$$

establishing perpendicularity.

Analogous to §2, the local construction of C proceeds by solving the differential equation

$$\frac{d\theta}{dr} = \frac{\tan x}{r},$$

where ψ and r are obtained from $\underline{C}$ and $\bar{C}$. To see that C is the desired curve, we work backward through the equations above to show that the arclength parameter on $\underline{C}$ agrees with that on C. Also, we need to calculate $d\psi + d\theta$, as before, for closure and convexity considerations.

$$\begin{aligned}\frac{d\psi}{dx} + \frac{d\theta}{dx} &= \frac{d}{dx}\left[\tan^{-1}\left(\frac{y-u}{z-x}\right) + \alpha + 90°\right] + \frac{\tan\psi}{r}\frac{dr}{dx}\\ &= \frac{\dfrac{d(y-u)}{dx}(z-x) - (y-u)\left(\dfrac{dz}{dx}-1\right)}{r^2} + \frac{d\alpha}{dx} + \frac{\tan\psi}{r}\frac{dr}{dx}\\ &= \frac{1}{r^2}\left[(z-x)\left(\tan(\alpha-\psi) + \frac{\cos\alpha}{\cos\psi}\frac{dr}{dx}\right) - (y-u)\left(\frac{\sin\alpha}{\cos\psi}\frac{dr}{dx} - 1\right) + r^2\frac{d\alpha}{dx} + r\tan\psi\frac{dr}{dx}\right]\\ &= \frac{1}{r^2}\frac{dr}{dx}\left[(z-x)\frac{\cos\alpha}{\cos\psi} - (y-u)\frac{\sin\alpha}{\cos\psi} + r\tan\psi + r^2\frac{d\alpha}{dr}\right] + \frac{1}{r^2}\left[\frac{(z-x)^2}{y-u} + y - u\right]\\ &= \frac{1}{y-u} + \frac{1}{r}\frac{dr}{dx}\left[\frac{\sin(\alpha-\psi)\cos\alpha}{\cos\psi} - \frac{\cos(\alpha-\psi)\sin\alpha}{\cos\psi} + \tan\psi + r\frac{d\alpha}{dr}\right]\\ &= \frac{1}{y-u} + \frac{1}{r}\frac{dr}{dx}\left[r\frac{d\alpha}{dr}\right]\\ &= \frac{1}{y-u} + \frac{d\alpha}{dx}.\end{aligned}$$

Hence C is convex if and only if

$$\frac{1}{y-u} + \frac{d\alpha}{ds} \geqslant 0.$$

The integral condition for closure is

$$\int \frac{1}{y-u}\,dx = 2\pi - \Delta\alpha$$

where the integral and the change in α are taken over one period. Note that y and u are the "heights" of the curves $\bar{C}$ and $\underline{C}$, but not necessarily above the same point. A point (x,y) lies above (z,u) only when $\alpha - \psi = 90°$ or, equivalently, the normal to (x,y) is vertical.

Previous considerations carry through with the obvious changes. Convexity corresponds to the curvature of $C(s)$ being greater than the curvature of $\underline{C}(s)$. Flat sides correspond to congruent pieces of C and $\underline{C}$ which roll against each other.

Finally one can ask the general question: Given $\bar{C}$, is there a pair $(C,\underline{C})$ of curves so that C rolls on $\underline{C}$ to produce $\bar{C}$. If no smoothness is required, the answer is yes, but some curves $\bar{C}$ admit no pairs $C,\underline{C}$ with the curvature of $C(s)$ greater than the curvature of $\underline{C}(s)$, C, $\underline{C}$ smooth. Details are left to the reader.

References

1. I. I. Artobolevskii, Mechanisms for the Generation of Plane Curves, Pergamon Press, New York, 1964.
2. W. H. Besant, Roulettes and Glissetes, Deighton, Bell, London, 1870.
3. R. C. Yates, A Handbook on Curves and Their Properties, 2nd ed., Edwards, Ann Arbor, Mich., 1942.

STOCHASTIC INDEPENDENCE AND SPACE-FILLING CURVES

JOHN A. R. HOLBROOK

Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada

1. Introduction. This note deals with a rather surprising relationship between "space-filling curves" and the elementary notions of probability theory. A representative example of this relationship may be stated briefly as follows: Properly constructed Peano curves may be viewed as sequences of continuous (but nonconstant) stochastically independent functions on the probability space [0,1] (see Theorems 2 and 7, and Corollary 5, below). We have used these ideas in the classroom for a number of years, and since casual discussions with several probabilists have suggested that such relationships are not widely known we offer the present exposition. We have learned from [**1**], however, that ideas related to some of those presented below occurred in the work of A. M. Garsia prior to 1974 (see Remark 3, below, for more details). It appears from [**1**] that Garsia's motivation was quite different from ours. Nevertheless, a parallel evolution of ideas may have been at work, since we first heard the question of the existence of continuous independent functions over [0,1] raised (but not resolved) during a discussion after Garsia's probability class at Caltech, many years ago.

2. Independence and Continuity. The concept of independent events occurs inevitably at the very beginning of any study of probability. We recall that the technical definition that reflects this intuitive concept is the following: events A, B of a probability space (Ω,P) are *independent* provided $P(A \cap B) = P(A)P(B)$. Likewise one encounters at an early stage the concept of

John Holbrook studied mathematics at Queen's University in Kingston, Ontario, and at the California Institute of Technology, where, as a student of W. A. J. Luxemburg, he obtained a doctorate in 1965. He has taught at Caltech, at the UC campuses in La Jolla and Santa Cruz, at the University of Guelph, at the Universidad Simón Bolívar in Caracas, Venezuela, and in the graduate program of IVIC (Instituto Venezolano de Investigaciónes Científicas). Although his published research is mainly in functional analysis and operator theory, he believes that good scientific health requires excursions into a wide variety of mathematical areas. The present article may illustrate that philosophy.—*Editors*

independent *random variables*, i.e., (real) functions f, g on Ω whose values are determined by independent events. Here $f^{-1}(J)$, $g^{-1}(K)$ are required to be independent events for any choice of real intervals J, K. Moreover, it is often possible to represent a probabilistic phenomenon by means of the most easily visualized probability space: $\Omega = I = [0,1]$, with the probability $P(A)$ of an event $A \subset I$ defined as the "length" or Lebesgue measure $|A|$ of A. For example, consider the well-known Rademacher functions $R_1, R_2, \ldots$, defined by the relations $R_n(t) = (-1)^k$ for $t \in [k2^{-n},(k+1)2^{-n})$ (here k may have the values $0,1,\ldots,2^n - 1$; for completeness we should also require $R_n(1) = 1$). These functions may be said to represent the experiment of making successive (independent) tosses of a (fair) coin, where we interpret $R_n = +1$ as "heads on the nth toss" and $R_n = -1$ as "tails on the nth toss." Thus the model of this experiment is the following: pick a point t "at random" in $I = [0,1]$; then the sequence

$$R_1(t), R_2(t), R_3(t), \ldots$$

determines the outcome (sequence of heads and tails). One can easily check that the functions $R_1, R_2, \ldots$ are mutually independent over I in the technical sense:

$$\left|\{t\colon R_{n_1}(t) = \varepsilon_1, R_{n_2}(t) = \varepsilon_2, \ldots, R_{n_m}(t) = \varepsilon_m\}\right| = (2^{-m} =) \prod_{i=1}^{m} \left|\{t\colon R_{n_i}(t) = \varepsilon_i\}\right|,$$

for all choices of $n_1 < n_2 < \cdots < n_m$ and $\varepsilon_i = \pm 1$.

Now the Rademacher functions are decidedly "jumpy," and it is natural to wonder whether *continuous* functions f, g on I can be stochastically independent. Of course, when we turn to "higher-dimensional" probability spaces, it is easy to give natural examples of continuous, independent functions. One such example is the natural model for the experiment of choosing a point at random in a square. Let $\Omega = I \times I, P(A) =$ the area (2-dimensional Lebesgue measure) of A; then the coordinate functions $f(x,y) = x$ and $g(x,y) = y$ are continuous and independent over Ω. A little reflection reveals, however, that continuous independent functions f, g over I must be wild indeed (we exclude the trivial case where one function is constant). The following simple proposition shows just how wild f and g must be and, at the same time, points the way to the construction of examples.

PROPOSITION 1. *Suppose f, g are real functions on $I = [0,1]$ that are stochastically independent (with respect to Lebesgue measure). Suppose also that f and g are continuous but not constant. Then $\gamma(t) = (f(t), g(t))$ defines a space- (area-) filling curve; indeed,* $\gamma(I) = f(I) \times g(I)$.

Proof. Since f and g are continuous but not constant, $f(I)$ and $g(I)$ are compact intervals of positive length. Consider $(x,y) \in f(I) \times g(I)$. Given $\varepsilon > 0$, let $J = (x - \varepsilon, x + \varepsilon)$. Since f is continuous $|f^{-1}(J)| > 0$. Similarly, if $K = (y - \varepsilon, y + \varepsilon)$, $|g^{-1}(K)| > 0$. By independence,

$$\left|f^{-1}(J) \cap g^{-1}(K)\right| = \left|f^{-1}(J)\right| \cdot \left|g^{-1}(K)\right| \qquad (> 0),$$

so that there is some $t \in f^{-1}(J) \cap g^{-1}(K)$. For such t, $\|\gamma(t) - (x,y)\| \leqslant \sqrt{2}\varepsilon$. Clearly, then, $\gamma(I)$ is dense in $f(I) \times g(I)$ and since $\gamma(I)$ must also be compact, we have $\gamma(I) = f(I) \times g(I)$.

Q.E.D.

Since space-filling curves have been familiar, if bewildering, mathematical objects since their invention by Peano, the obvious next step is to examine the construction of such curves to see if independence of the coordinate functions can, in fact, be arranged. There results the following theorem.

THEOREM 2. *Functions of the type described in Proposition 1 exist. In fact, there are continuous, independent functions f, g on I such that each is uniformly distributed over I (i.e., $|f^{-1}(J)| = |g^{-1}(J)| = |J|$ for each subinterval J of I).*

We omit the detailed proof of this theorem since we are going to prove more general results in the next section (see Theorem 4 and Corollary 5). The basic idea, however, can be explained

immediately to anyone familiar with the classic constructions of Peano curves $\gamma: I \to I \times I$. In Hilbert's dyadic construction, for example, γ emerges as the uniform limit of curves $\{\gamma_n\}_1^\infty$ where γ_n traces in some order through the 4^n subsquares of the regular partition A_n of $I \times I$ into subsquares having dimensions 2^{-n} by 2^{-n}. If γ_n is parametrized appropriately, it will "spend an equal time" in each of the subsquares, that is:

$$\left|\{t: \gamma_n(t) \in S\}\right| = 4^{-n} \qquad (S \in A_n). \tag{1}$$

Now suppose that $\gamma_n = (f_n, g_n)$ and that $X, Y \subset I$ are (disjoint) unions of dyadic intervals of order n:

$$X = \bigcup_{a=1}^{p} J_a, \qquad Y = \bigcup_{b=1}^{q} K_b.$$

Then

$$|f_n^{-1}(X) \cap g_n^{-1}(Y)| = |\{t: \gamma_n(t) \in X \times Y\}| = \sum_{a,b} \left|\{t: \gamma_n(t) \in J_a \times K_b\}\right|.$$

There are pq terms in this sum and each $J_a \times K_b \in A_n$ so that, in view of (1), we have

$$\left|f_n^{-1}(X) \cap g_n^{-1}(Y)\right| = pq4^{-n} = (p2^{-n})(q2^{-n}). \tag{2}$$

In particular, if we let $q = 2^n$ so that $Y = I$, (2) tells us that $|f_n^{-1}(X)| = p2^{-n}$, and similarly $|g_n^{-1}(Y)| = q2^{-n}$. It follows from (2) that

$$\left|f_n^{-1}(X) \cap g_n^{-1}(Y)\right| = \left|f_n^{-1}(X)\right| \cdot \left|g_n^{-1}(Y)\right|$$

for all sets X, Y of the form described above. Thus it is easy to believe that the limit functions $f = \lim f_n$, $g = \lim g_n$ should be independent. Since $|f_n^{-1}(X)| = p2^{-n} = |X|$, we may also expect that f is uniformly distributed, and the same argument applies to g.

REMARK 3. To expand upon our comments in the introduction, there appears to be a close connection between Theorem 2 and the work of A. M. Garsia [**1**]. In [**1**, 2.24] Garsia describes the construction of a measure-preserving map γ of I onto $I \times I$ such that γ is also Lipschitzian of order $\frac{1}{2}$ (and hence continuous). In probabilistic terms this evidently means that the coordinate functions f, g of γ are independent and each has the uniform joint distribution over I, i.e., the conditions of Theorem 2 are met. In the next section we shall examine results similar to Theorem 2 but involving more general joint distributions.

3. Joint Distributions. For our purpose, we may take the joint distribution of a pair of (Borel-measurable) functions $f, g: I \to I$ to be the measure $P_{f,g}$ defined for each Borel subset Z of $I \times I$ by

$$P_{f,g}(Z) = \left|\{t: (f(t), g(t)) \in Z\}\right|.$$

The discussion of Section 2 justifies interest in the following general question: Which joint distributions can be obtained from pairs f, g of *continuous* functions from I to I? Our next theorem provides a partial answer.

THEOREM 4. *Let P be a (nonnegative) Borel measure on $I \times I$ with $P(I \times I) = 1$ and with "strictly positive density" in the following sense:*

$$\inf\{P(S)/area(S): S \text{ a subsquare of } I \times I\} = c > 0. \tag{3}$$

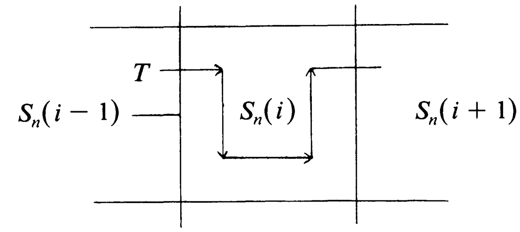
Then there exist continuous functions $f, g: I \to I$ such that their joint distribution $P_{f,g} = P$.

Proof. Let A_n denote the regular partition of $I \times I$ into the 4^n disjoint subsquares of the form $J \times K$ where J and K are intervals from the list

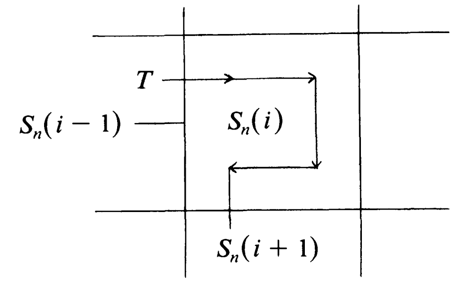
$$[0,2^{-n}],(2^{-n},2\cdot 2^{-n}],(2\cdot 2^{-n},3\cdot 2^{-n}],\ldots.$$

Note that each $S \in A_n$ is the union of four elements of A_{n+1}. We shall use one of the standard constructions of a space-filling curve to arrange the elements of each A_n in order. A variant of the "dyadic" construction due to Hilbert, rather than the "ternary" method used originally by Peano (see H. C. Kennedy [2]), will be convenient. We shall give the details of this construction in a form suitable for our application. Suppose that the elements of A_n have been listed as $S_n(1), S_n(2), \ldots, S_n(4^n)$ in such a way that $S_n(i+1)$ is adjacent to $S_n(i)$ $(1 \leq i < 4^n)$, i.e., the distance between the center $C_n(i)$ of $S_n(i)$ and $C_n(i+1)$ is 2^{-n}. Evidently this can be done for $n = 1$. The essential step is to observe that the same can then be done for A_{n+1} so that in addition

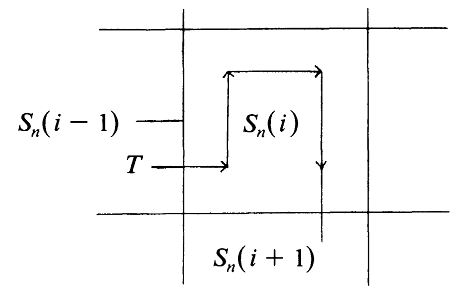
$$S_n(i) = S_{n+1}(4i-3) \cup S_{n+1}(4i-2) \cup S_{n+1}(4i-1) \cup S_{n+1}(4i). \tag{4}$$



(5)



(6)



(7)

Except for rotations and reflections, there are just three cases to consider, and these are handled in diagrams (5), (6), and (7). In these diagrams $S_n(i)$ is situated in the various possible ways with respect to $S_n(i-1)$ and $S_n(i+1)$, and with respect to $T = S_{n+1}(4(i-1))$. We are assuming that the choice of $S_{n+1}(1), S_{n+1}(2), \ldots, S_{n+1}(4(i-1))$ has been successfully done. The arrows indicate the appropriate choices for $S_{n+1}(4i-3)$, $S_{n+1}(4i-2)$, $S_{n+1}(4i-1)$, and $S_{n+1}(4i)$. Following these prescriptions inductively we obtain lists $\{S_n(i)\}$ of the elements of A_n for all n such that (4) is satisfied along with

$$\|C_n(i) - C_n(i+1)\| = 2^{-n} \qquad (1 \leq i < 4^n). \tag{8}$$

We now define the intervals $\{I_n(i)\}_{i=1}^{4^n}$ as follows:

$$I_n(1) = [0, P(S_n(1))],$$

$$I_n(i) = \left(\sum_{j<i} P(S_n(j)), \sum_{j\leqslant i} P(S_n(j)) \right] \qquad (1 < i \leqslant 4^n). \tag{9}$$

Since $\Sigma_{j\leqslant 4^n} P(S_n(j)) = P(I \times I) = 1$, we see that $I_n(1),\ldots,I_n(4^n)$ is a partition of I. Hence we may define functions γ_n: $I \to I \times I$ by requiring that $\gamma_n(t) = C_n(i)$ for all $t \in I_n(i)$ $(1 \leqslant i \leqslant 4^n)$. From (4) we see that

$$P(S_n(i)) = \sum_{j=4i-3}^{4i} P(S_{n+1}(j))$$

so that, in view of (9),

$$I_n(i) = I_{n+1}(4i-3) \cup I_{n+1}(4i-2) \cup I_{n+1}(4i-1) \cup I_{n+1}(4i). \tag{10}$$

Now if $4i - 3 \leqslant j \leqslant 4i$, the coordinates of $C_{n+1}(j)$ differ from those of $C_n(i)$ by $2^{-(n+2)}$ so that, if f_n, g_n denote the coordinate functions of γ_n, we see from (10) that for all $t \in I$

$$|f_{n+1}(t) - f_n(t)| = 2^{-(n+2)} \quad \text{and} \quad |g_{n+1}(t) - g_n(t)| = 2^{-(n+2)}. \tag{11}$$

Clearly, then, the sequences $\{f_n\}$, $\{g_n\}$ converge (uniformly) to functions f,g from I to I. It remains to show that f and g are continuous (although the f_n and g_n are not) and that $P_{f,g} = P$.

Now, since by construction $|I_n(i)| = P(S_n(i))$, it is clear from (3) that $|I_n(i)| \geqslant c4^{-n}$. Hence if $|t - s| < c4^{-n}$, t and s lie either in the same $I_n(i)$ or in neighboring intervals $I_n(i), I_n(i+1)$, so that, by (8), $|f_n(t) - f_n(s)|$ and $|g_n(t) - g_n(s)|$ are no greater than 2^{-n}. From (11) we evidently have

$$|f(t) - f_n(t)| \leqslant \sum_{n+2}^{\infty} 2^{-k} = 2^{-(n+1)}.$$

It follows that when $|t - s| < c4^{-n}$

$$\begin{aligned} |f(t) - f(s)| &\leqslant |f(t) - f_n(t)| + |f_n(t) - f_n(s)| + |f_n(s) - f(s)| \\ &\leqslant 2^{-(n+1)} + 2^{-n} + 2^{-(n+1)} = 2^{-(n-1)}. \end{aligned}$$

Certainly, then, f must be continuous, and the continuity of g is equally clear.

Consider a square $S = S_n(k)$ in A_n and let $J = I_n(k)$. Let γ denote the curve to which the sequence $\{\gamma_n\}$ converges (uniformly over I). We first show that $|J - \gamma^{-1}(S)| = 0$. For each $m \geqslant n$ let

$$H_m = \{ i \colon (J - \gamma^{-1}(S)) \cap I_m(i) \neq \varnothing \}.$$

Clearly the sequences of sets

$$E_m = \bigcup_{i \in H_m} I_m(i) \qquad (m = n, n+1, \ldots)$$

and

$$F_m = \bigcup_{i \in H_m} S_m(i) \qquad (m = n, n+1, \ldots)$$

are monotone decreasing. Hence, if $F = \cap_m F_m$, we have $P(F) = \lim P(F_m)$. On the other hand, for each m,

$$\left| J - \gamma^{-1}(S) \right| \leqslant \sum_{i \in H_m} \left| I_m(i) \right| = \sum_{i \in H_m} P(S_m(i)) = P(F_m),$$

so that it suffices to show that $P(F) = 0$. Now if $Q \in F$, then for each $m (\geqslant n)$ there is some $i_m \in H_m$ such that $Q \in S_m(i_m)$. Suppose that

$$t_m \in (J - \gamma^{-1}(S)) \cap I_m(i_m).$$

We observe that $\|\gamma(t_m) - \gamma_m(t_m)\| \to_m 0$ and that $\gamma_m(t_m) \in S_m(i_m)$. It follows that $\|\gamma(t_m) - Q\| \to_m 0$. Evidently, then,

$$F \subset \overline{\gamma\big(J - \gamma^{-1}(S)\big)}. \tag{12}$$

On the other hand, $\gamma(J) \subset \bar{S}$ since $\gamma_m(J) \subset S$ for each $m \geqslant n$. Since $\gamma(J - \gamma^{-1}(S)) \subset S'$ (the complement of S) we conclude from (12) that

$$F \subset \overline{(\bar{S} \cap S')}.$$

However, it is easy to check that the definition of A_n ensures that, for each $S \in A_n$, $\bar{S} \cap S'$ is closed. Thus $F \subset \bar{S} \cap S'$, and, since F is certainly a subset of S, we conclude that $F = \varnothing$ so that, indeed, $P(F) = 0$.

Finally, we note that, since $|J - \gamma^{-1}(S)| = 0$, we have

$$P_{f,g}(S) = |\gamma^{-1}(S)| \geqslant |J| = P(S).$$

Since this is true for all $S \in A_n$ and $P_{f,g}(I \times I) = P(I \times I) = 1$, we see that $P_{f,g} = P$ on each A_n so that the class of Borel sets

$$\mathfrak{M} = \{B\colon B \text{ is Borel and } P_{f,g}(B) = P(B)\}$$

includes the field of sets generated by $\cup_n A_n$. By the well-known theorem on monotone classes, $\mathfrak{M} =$ all Borel sets (in $I \times I$), i.e., $P_{f,g} = P$. Q.E.D.

COROLLARY 5. *Suppose that $a(x)$ and $b(y)$ are continuous positive functions on I such that $\int_0^1 a(x)\,dx = \int_0^1 b(y)\,dy = 1$. Then there exist continuous functions f,g: $I \to I$ such that the distribution function of f has density $a(x)$, the distribution function of g has density $b(y)$, and f,g are independent over I.*

Proof. What is required here is a pair f,g with joint density function $a(x)b(y)$, i.e., such that $P_{f,g} = P$ where

$$P(Z) = \iint_Z a(x)b(y)\,dx\,dy.$$

Since (3) is clearly satisfied with

$$c = \Big(\min_I a(x)\Big)\Big(\min_I b(y)\Big),$$

the existence of such (continuous) f,g follows from Theorem 4. Q.E.D.

4. Sequences of Continuous Random Variables. It is not hard to see how the proof of Theorem 4 may be extended to yield the following result.

THEOREM 6. *Let d be any positive integer and let P be a (nonnegative) Borel measure on*

$$I \times I \times \cdots \times I = I^d$$

with $P(I^d) = 1$. Suppose that $\inf\{P(S)/(d-\text{volume of } S)$: *$S$ is a d-dimensional subcube of $I^d\}$ is positive. Then there exist continuous functions f_k: $I \to I$ $(k = 1,2,\ldots,d)$ such that the joint distribution measure $P_{f_1,f_2,\ldots,f_d} = P$.*

It may also be of interest to note that in the uniformly distributed case an infinite sequence of independent, continuous random variables on I can be constructed directly from Theorem 2.

THEOREM 7. *Suppose that f,g: $I \to I$ are uniformly distributed over I, continuous on I, and stochastically independent. Then the functions*

$$h_n = f \circ \underbrace{g \circ g \circ \cdots \circ g}_{(n-1)\ \text{times}} \qquad (n = 1,2,\ldots)$$

(*where* "$\circ$" *denotes functional composition*) *are continuous, uniformly distributed in* I, *and mutually independent over* I.

Proof. We must verify that the joint distribution of $h_1,\ldots,h_d$ is uniform in I^d $(d=1,2,\ldots)$, i.e., that, for all subintervals $J_1,\ldots,J_d$ of I,

$$\left|\{t\colon h_1(t)\in J_1,\ldots,h_d(t)\in J_d\}\right| = |J_1|\cdots|J_d|. \tag{13}$$

This is easy to see by induction on d. Since (13) is clear when $d=1$, we assume (13) for some $d\geqslant 1$ and compute as follows:

$$\left|\{t\colon h_1(t)\in J_1,\ldots,h_{d+1}(t)\in J_{d+1}\}\right| = \left|\{t\colon f(t)\in J_1, g(t)\in K\}\right|$$

where

$$K=\{s\colon f(s)\in J_2,(f\circ g)(s)\in J_3,\ldots,(f\circ g\circ\cdots\circ g)(s)\in J_{d+1}\}.$$

Hence, by the assumptions on f,g we have:

$$\begin{aligned}\left|\{t\colon h_1(t)\in J_1,\ldots,h_{d+1}(t)\in J_{d+1}\}\right| &= |J_1|\cdot|K|\\ &= |J_1|\cdot\left|\{s\colon h_1(s)\in J_2,\ldots,h_d(s)\in J_{d+1}\}\right|\\ &= |J_1|\cdot|J_2|\cdots|J_{d+1}|,\end{aligned}$$

using the inductive hypothesis. Q.E.D.

REMARK 8. We have just seen that there exists a sequence of continuous (but nonconstant) mutually independent functions on I. We remark that no collection of such functions is more than countable since if $\{f_\alpha\}_{\alpha\in A}$ is such a collection, it is easy to see that the functions $\{g_\alpha\}_{\alpha\in A}$ are orthonormal in the (separable) space $L^2(I)$, where

$$g_\alpha(t)=\left(h_\alpha(t)-\int_0^1 h_\alpha(s)\,ds\right)\Bigg/\left(\int_0^1 h_\alpha^2(s)\,ds-\left(\int_0^1 h_\alpha(s)\,ds\right)^2\right)^{1/2}.$$

REMARK 9. By an obvious extension of Proposition 1, the functions $h_1,\ldots,h_d$ constructed in Theorem 7 define, for each d, a d-dimensional (continuous) space-filling curve I onto I^d.

The work for this paper was supported in part by the National Research Council of Canada under grant #A8745.

References

1. A. M. Garsia, Combinatorial inequalities and smoothness of functions, Bull. Amer. Math. Soc., 82 (1976) 157–170.
2. H. C. Kennedy, Selected Works of Giuseppe Peano, University of Toronto Press, 1973.

TWENTY YEARS LATER: HIGH SCHOOL STUDENTS WHO SHOWED PROMISE IN MATHEMATICS

NURA D. TURNER
Department of Mathematics, State University of New York, 1400 Washington Ave., Albany, NY 12222

Before we had a U.S.A. Mathematical Olympiad, we had only one national measuring stick for pinpointing students demonstrating keen ability in mathematics: a competition which has become

known as the Annual High School Mathematics Examination. That examination has been sponsored by the Mathematical Association of America (MAA) since 1957 (predating Sputnik).

Beginning with the first competition in 1958, I served as the chairman of the Upstate New York Contest Section (now the Seaway Section) for thirteen years. Since the contest consumed a great deal of time and effort of some MAA members, as well as what I then estimated as at least half a million dollars a year, I felt that it would be worthwhile to keep track of the future progress of the high achievers from the first years of the test. I selected, for the first three years of the competition, those students in the Upstate New York Section who scored in the top one percent (31 students for 1958, 34 for 1959, 26 for 1960). In addition, I chose 27 more students who, in the national 1958 competition, scored in the top .03 percent. For twenty years, over a hundred of these 118 students have faithfully answered my requests for current information on their academic work in college and graduate school, their careers and families.

Some statistics are interesting.

41.5 percent entered college intending to earn a bachelor's degree in mathematics.
94.1 percent earned a bachelor's degree in some field; 5.9 percent completed this in three years.
37.3 percent earned bachelor's degrees in mathematics.
42.4 percent earned a Ph.D. or M.D.; 32 of these were "Upstaters"; the remainder were from the highly selective national group.
15.3 percent earned a doctorate in mathematics.
55.9 percent have pursued careers in mathematically related fields and mathematics.

In 1966, the three main career interests (in order of choice) were mathematics, engineering, and physics. By 1978, the established careers most frequently cited were mathematics, engineering, and computer science (in that order).

Twelve mathematicians are on the faculties of universities in the United States. Two of those rose especially fast on the academic ladder: one at the University of California became an associate professor four years after earning his Ph.D., and full professor five years later, and another at an eastern university became a full professor in even less time.

Only a limited amount of mathematics has been used by the eighteen members with positions in industry and the federal government (these eighteen have at least a bachelor's degree in mathematics).

Since some of those who planned to major in mathematics in college left that field (or wish they had left) before earning a bachelor's degree, it might be useful to quote some of their criticisms.

> Through my junior year I always had a deep interest in math per se—and found playing around with problems an enjoyable diversion. However, pure mathematics became too abstract for my less than rigorous personality and I went to Stanford for a MBA in order to head further in the direction of applying my background to real, dynamic situations.

> I had to choose between math and physics. I chose physics because I don't have the patience to prove my results with the care that mathematicians demand. There is no use giving a complete, axiomatic proof of anything because no one will really believe it until he has done the experiment anyway.

> My decision to change my major in college from math to English came in a rather strange instance. I was taking one math course at the time—my junior year. One evening I stayed out late. The next morning my alarm clock jarred me awake. There was still time to get to my math class, but I shoved the clock away and went back to sleep. Actually, this action only made final a decision I had been planning for some time. Though I had been getting the feeling that math was just a game, having little relevance to everyday life, there was one other important consideration; I liked the academic life, and wanted to teach. I dreaded the boredom of teaching the same courses for years. In English there is more opportunity for a variety of approach, and unlike math, every student paper is unique. There is more of a chance to deal with the individual.

> A discouraging and unsuccessful bout with the calculus my freshman year helped to direct me away from the

scientific and engineering fields. Whether I would have become interested in the liberal arts and especially in philosophy if I had not conceived a dislike for math is questionable. My course in symbolic and mathematical logic, taught by a superb teacher, reawakened my appreciation for and delight in mathematical thought. I wish now that I had persevered in mathematical studies.

Although I did well in math in high school with an overwhelming extracurricular interest and was able to start college math with advanced calculus, I lost interest in the more abstract aspects, never mastered abstract algebra, and by my sophomore year was out of math looking for a different field. I dabbled in physics and mathematical linguistics but eventually became intrigued with biology. My medical specialty is psychiatry and much of my recent work has been on drug abuse. Recently I met a younger fellow who grew up in the math genius culture of Stuyvesant High School at the time LSD was "the way" and it was very strange to hear of math students doing very well while heavily into drugs. Usually LSD, etc., are thought to disorganize abstract mental processes. The super abstract mathematics pursuit is a "trip" in itself.

Even those who persevered in their study of mathematics to complete the doctorate, or who re-entered the field after having left it, were not immune to expressing similar criticism.

My enjoyable experiences in mathematics have been almost entirely outside the classroom.

At no time, even in special classes or in mathematics team-practice sessions, did any teacher indicate what mathematicians do or what mathematics is all about. In my first two years of college I was not enlightened further. As a junior my mathematics adviser merely asked which courses I was interested in and signed her approval. Non-science majors on the other hand had an eye-opening introductory course on the foundations of mathematics. As an upper junior, I registered for an advanced version of this course and finally learned what mathematicians call math. I would say that the educational system wasted about ten years of my mathematical life.

My option to pursue physics probably resulted from having an exciting physics instructor and a very boring and bored mathematics one during my first year in college. [I mention this here because interest in mathematics has been reawakened in this young man to the extent that he is a member of the Department of Applied Mathematics at MIT.]

I was a junior before I took a course from a real live mathematician and found out what things were all about in the field. There just aren't enough people around who understand what is involved in being a professional mathematician, and most of them are pretty busy being just that, with little time left for giving advice to younger people. The attitude here seems to be if someone has the ability, he'll see what's happening all by himself.

Yet those who earned at least a bachelor's degree in mathematics have also commented on the positive influence of mathematics in their careers.

I could not be where I am today without a solid background in mathematics. I have had experience in developing the guidance program used to get the Saturn vehicles into orbit with a minimum of fuel usage, which required knowledge of the calculus of variations to understand the theory. At the same time, I developed models and simulations to verify the resulting computer programs, which made extensive use of most undergraduate mathematics.

Clearly a career in theoretical physics would be out of the question if one were not willing to learn some fairly sophisticated mathematics. Mathematics is the only way one can express and manipulate physical ideas. Inspiration may come from other quarters, but for the demonstration and elaboration of an idea, mathematics is indispensable.

Even as a physicist, what success I've had owes more to skills with a wrench learned in fixing farm machinery than to any talent in mathematics. But if I had the chance and could do something I haven't done, I would like to be able to think mathematics.

I believe that the courses I took in math helped develop my analytical abilities and may have contributed to my great success in law school.

I am sure my partial predilection for mathematics guided me towards linguistics and helped me in it. Linguistics has felt the influence of probability and is described as an effort to mathematize language.

My interest in mathematics has led me to pursue research in a medical area, which allows me to use my

mathematical training and interest. I have been involved in hematological research which basically has been theoretical mathematics with strong reliance upon computer techniques.

I feel that the study of mathematics helped me develop a crisp analytic approach to legal problems. My free lance study of symbolic logic was of direct use in law school: a logical formulation of the problem is not a solution, but certainly is a healthy start.

While I left the formal study of math after my second undergraduate year, I never departed from the benefits of such study. I am making practical use of my math training running a business, managing large sums, keeping records et al.... but nothing like the fun and challenge of real abstractions.

It has been stated that those working in industry and with the federal government have used only a limited amount of mathematics. It includes combinatorics and algebra in work with cryptologic problems; statistical analysis of problems dealing, for example, with how many lives are saved by seat belts; statistical analysis in budgeting; computer graphics in work with computer software; logic concepts for formally describing computer programs; optimization theory in military design; operations research in management science and information systems; development of algorithms in solving problems such as the presenting of information on display terminal screens; and in the use of algorithms, the employment of techniques from discrete and combinatorial mathematics, matrix algebra, and complexity theory, such as the use of binary trees, hashing techniques, and tricks of ordering computations so as to minimize the number of operations required.

What can we deduce from our data?

The Annual High School Mathematics Examination of at least 1958, 1959, and 1960 did a good job of identifying high school students showing promise in mathematics. The majority of those students not only started out to prepare for careers in mathematics or in mathematics-related fields but established careers in those areas. Moreover, the students were ambitious, set their goals high, and went straight ahead toward attaining them. High School teaching was not chosen as a career by these students with exceptional mathematical ability.

Our educational system, grades K through 12, needs to identify early those students with marked ability in mathematics and then to provide the opportunity for their having special training.

The comments I have quoted, and others, reaffirm the main conclusions of my report in 1967 [**1**]:

1. Not enough is being done to introduce students to a general knowledge of what mathematics is about.
2. Mathematics counseling on the undergraduate level is poor.
3. Mathematics is poorly taught at the calculus level, both elementary and advanced.
4. We need to strengthen the Contest program by using it to improve the educational process. One of the study members remarked that when he was in high school the "Contest catered to people who were trained in New York City math teams who learned through drill how to solve certain types of mathematical problems." He added that he hoped the U.S.A. Mathematical Olympiad has improved the situation.

Indeed, I wonder about the winners of the U.S.A. Mathematical Olympiad... their future in college and graduate school, and their careers. I have already begun a similar study on their academic and career progress. It would be valuable to continue further with them to learn if their children show the same marked ability in mathematics.

Reference

1. Nura D. Turner, National aspects of the MAA-SA contest in the development of talent, this MONTHLY, 74 (1967) 60–67.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis , MO 63121.

HOMEOMORPHISMS OF CUBES INTO EUCLIDEAN SPACES

Y. STERNFELD
Department of Mathematics, University of Haifa, Haifa, Israel

By the classical result of Menger and Nöbeling every separable metric space of topological dimension n is homeomorphic to a subset of $(2n+1)$-dimensional Euclidean space. It is also known that the dimension $2n+1$ is the best possible, in the sense that for each $n \geqslant 1$ there exists an n-dimensional compact metric space X_n that does not embed homeomorphically in the $2n$-dimensional Euclidean space. (X_1 can be taken to be the union of edges of a tetrahedron and of a segment joining two disjoint edges (see Fig. 1). For higher dimensions, see [**1**, p. 63].)

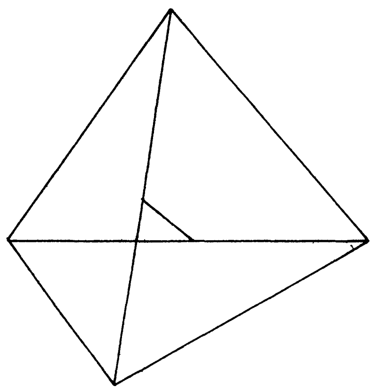


FIG. 1

In this note we wish to point out another aspect in which the dimension $2n+1$ is the best possible.

From the usual proof of the Menger-Nöbeling theorem (see [**1**, p. 56]) we actually obtain a stronger result than the one stated above. Restricting ourselves to compact spaces it reads as follows:

STRONGER VERSION OF THE MENGER-NÖBELING THEOREM. *If X is a compact metric space of dimension n, then the homeomorphisms in $C(X,R^{2n+1})$ form a dense G_δ set.*

Note that R^k is k-dimensional Euclidean space, and $C(X,R^k)$ is the space of continuous functions from X to R^k. We look upon both spaces as Banach spaces over the reals with the norms

$$\|(x_1,x_2,\dots,x_k)\| = \max_{1\leqslant i\leqslant k} |x_i|$$

in R^k and

$$|||\Phi||| = \sup_{x\in X} \|\Phi(x)\|$$

in $C(X,R^k)$. By homeomorphisms we mean homeomorphisms into, that is, one-to-one mappings.

It turns out that this stronger version of the Menger-Nöbeling theorem fails to hold even for

the simplest n-dimensional spaces—namely, n-dimensional cubes—if the dimension $2n+1$ is reduced to $2n$. More precisely we have the following:

THEOREM. *The homeomorphisms in $C(I^n,R^{2n})$ are nowhere dense.*

(I^n denotes the n-dimensional cube $[-1,1]^n$.)

It seems as if the homeomorphisms are nowhere dense in $C(X,R^{2n})$ for every compact n-dimensional metric space, but we did not find a proof for this; the fact, however, that, in the case $X=I^n$, or more generally when X contains a copy of I^n, they are nowhere dense illustrates the minimality of the dimension $2n+1$. From the category point of view one can say that almost all elements of $C(I^n,R^{2n+1})$ are homeomorphisms, while in $C(I^n,R^{2n})$ the set of homeomorphisms, though not empty, is meager.

Our proof of the theorem is almost elementary. Hurewicz and Wallman's book **[1]** is an excellent reference for the topics mentioned in this article. The following lemma is the heart of the proof.

LEMMA. *There exists a function Φ in $C(I^n,R^{2n})$ and a positive number δ such that if $\Psi \in C(I^n,R^{2n})$ and $|||\Phi-\Psi|||<\delta$ then Ψ is not one to one.*

Proof. Let J_1 and J_2 be two disjoint closed subintervals of $I=[-1,1]$; and let ϕ_i, $i=1,2$, be continuous real-valued functions on I such that ϕ_i maps J_i homeomorphically onto $[-1,1]$ and is zero on J_{3-i}. Let $\Phi \in C(I^n,R^{2n})$ be the function defined by

$$\Phi(x_1,x_2,\ldots,x_n)=(\phi_1(x_1),\phi_1(x_2),\ldots,\phi_1(x_n),\phi_2(x_1),\phi_2(x_2),\ldots,\phi_2(x_n)).$$

We shall see that Φ possesses the desired property for a suitable δ.

REMARK. For $n=1$ it is easy to graph the image of I under Φ in R^2. (See Fig. 2.)

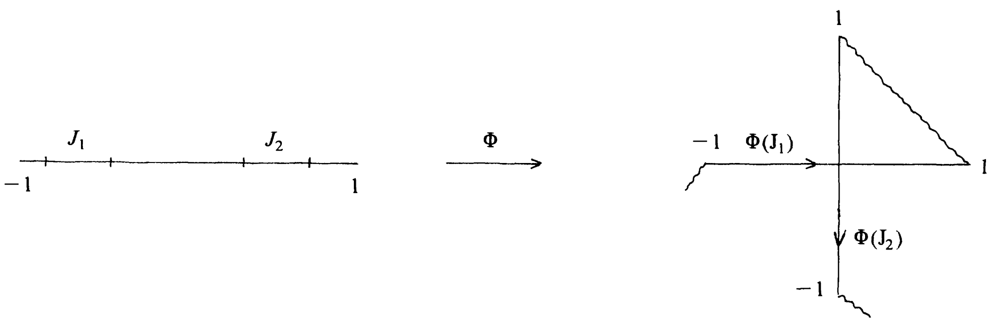


FIG. 2

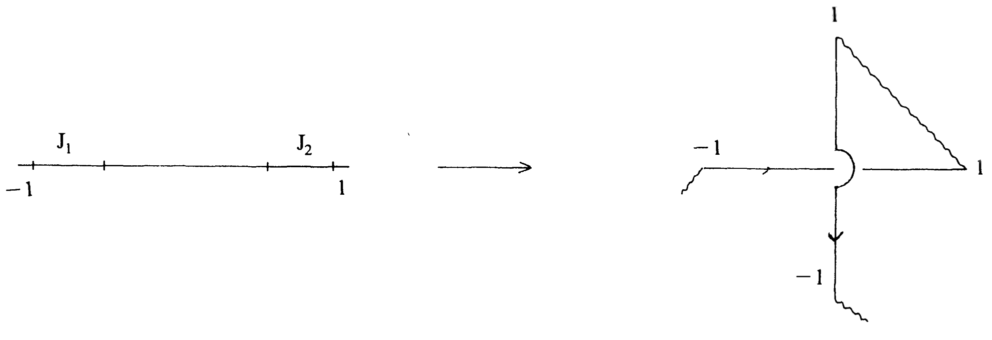


FIG. 3

It is intuitively evident from Fig. 2 that Φ cannot be closely approximated by a one-to-one mapping of I into R^2. Note, however, that if we allow mappings of I into R^3 then this can be easily done; the image of I under such a map is sketched in Fig. 3, illustrating the difference between $2n$ and $2n+1$. Set $D_i = J_i^n$, $i = 1,2$, and $D = D_1 \times D_2$. D_1 and D_2 are disjoint n-dimensional cubes in I^n and D is a $2n$-dimensional cube. If we realize R^{2n} as $R^n \times R^n$ then it follows that $\Phi(D_1) = I^n \times \{0\}$, $\Phi(D_2) = \{0\} \times I^n$ and hence $\Phi(D_1) \cap \Phi(D_2) = \{0\} \times \{0\}$.

For $\Psi \in C(I^n, R^{2n})$ consider the function $\overline{\Psi} \in C(D, R^{2n})$ defined by $\overline{\Psi}(u,v) = \Psi(u) - \Psi(v)$ (vector subtraction in R^{2n}), $(u,v) \in D = D_1 \times D_2$. It is easily checked that the operation $\Psi \to \overline{\Psi}$ is continuous, and that $\overline{\Phi}$ maps D homeomorphically onto $[-1,1]^{2n}$ and its boundary $\partial(D)$ onto $\partial([-1,1]^{2n})$. It follows that there exists a positive δ such that $|||\Phi - \Psi||| < \delta$ for some Ψ in $C(I^n, R^{2n})$ implies $(0,0,\dots,0) \notin \{(1-t)\overline{\Phi}(u,v) + t\overline{\Psi}(u,v) \colon 0 \leq t \leq 1, (u,v) \in \partial(D)\}$. (Actually one can take $\delta = \frac{1}{2}$; indeed, if $|||\Phi - \Psi||| < \frac{1}{2}$ then, for $(u,v) \in D$,

$$\begin{aligned}\|\overline{\Phi}(u,v) - \overline{\Psi}(u,v)\| &= \|\Phi(u) - \Phi(v) - (\Psi(u) - \Psi(v))\| \\ &\leq \|\Phi(u) - \Psi(u)\| + \|\Phi(v) - \Psi(v)\| < \tfrac{1}{2} + \tfrac{1}{2} = 1;\end{aligned}$$

hence, for $(u,v) \in \partial(D)$, $\overline{\Phi}(u,v)$ is in the boundary of $[-1,1]^{2n}$, and $\overline{\Psi}(u,v)$ is a vector in R^{2n} of distance < 1 from $\overline{\Phi}(u,v)$. It follows that the segment joining $\overline{\Phi}(u,v)$ and $\overline{\Psi}(u,v)$ will not pass through the origin.)

We claim now that if $|||\Phi - \Psi||| < \delta$ then Ψ is not one to one. Indeed, let $|||\Phi - \Psi||| < \delta$ and assume Ψ is one to one. Then $\Psi(D_1)$ and $\Psi(D_2)$ are disjoint; hence $(0,0,\dots,0) \notin \overline{\Psi}(D)$. Thus $\hat{\Psi}(u,v) = \overline{\Psi}(u,v)/\|\overline{\Psi}(u,v)\|$ is a well-defined element of $C(D, R^{2n})$, which maps D into the boundary of $[-1,1]^{2n}$. Moreover $\overline{\Phi}$ and $\overline{\Psi}$ are homotopic as mappings of $\partial(D)$ into $\partial([-1,1]^{2n})$, since by the choice of δ the deformation $F\colon \partial(D) \times [0,1] \to \partial([-1,1]^{2n})$, defined by

$$F(u,v,t) = \frac{(1-t)\overline{\Phi}(u,v) + t\overline{\Psi}(u,v)}{\|(1-t)\overline{\Phi}(u,v) + t\overline{\Psi}(u,v)\|},$$

is continuous, and it clearly satisfies $F(u,v,0) = \overline{\Phi}(u,v)$ and $F(u,v,1) = \hat{\Psi}(u,v)$.

Consider the mapping $\eta\colon D \to \partial(D)$ defined by $\eta(x) = \overline{\Phi}^{-1}(\hat{\Psi}(x))$, $x \in D$; η is well defined since $\overline{\Phi}$ is one to one, and it maps into $\partial(D)$ since $\hat{\Psi}$ maps into $\partial([-1,1]^{2n})$, and Φ^{-1} maps $\partial([-1,1]^{2n})$ onto $\partial(D)$. From the fact that $\hat{\Psi}$ and $\overline{\Phi}$ are homotopic on $\partial(D)$ it follows that $\eta = \overline{\Phi}^{-1} \circ \hat{\Psi}$ is homotopic to $\overline{\Phi}^{-1} \circ \overline{\Phi} =$ identity on $\partial(D)$. So η is a mapping of D into $\partial(D)$ that is homotopic to the identity on $\partial(D)$. Applying the extension theorem of Borsuk ([**1**, p. 86]) we can extend the identity mapping of $\partial(D)$ to the whole of D and obtain a continuous function $f\colon D \to \partial(D)$ that is the identity on $\partial(D)$, which is well known to be impossible ([**1**, p. 40). This proves the lemma.

Proof of the theorem. We wish to show that every open subset U of $C(I^n, R^{2n})$ contains a further open set V that is free of homeomorphisms. Let $U \subset C(I^n, R^{2n})$ be open, fix some $f \in U$, and let $r > 0$ be such that $|||f - g||| < r$ implies $g \in U$. We may assume without loss of generality that 0 ($=$ the origin in R^{2n}) is in the range of f. Let $x_0 \in I^n$ be a point with $f(x_0) = 0$. By the continuity of f there exists an n-cube $C_1 \subset I^n$ with $x_0 \in C_1$ and $|||f|||_{C_1} < r/2$. Let C_2 be another n-cube contained in the interior of C_1. By the lemma we can find an element Φ of $C(C_2, R^{2n})$ with $|||\Phi|||_{C_2} < r/2$ and $\delta > 0$ such that if $|||\Phi - \Psi|||_{C_2} < \delta$ then Ψ is not one to one (on C_2). Let $g \in C(I^n, R^{2n})$ be defined as follows: first, define $g_1\colon C_2 \cup \partial(C_1) \to R^{2n}$ by

$$g_1(x) = \begin{cases} \Phi(x) & \text{if } x \in C_2 \\ f(x) & \text{if } x \in \partial(C_1). \end{cases}$$

Then $|||g_1|||_{C_2 \cup \partial(C_1)} < r/2$. Applying Tietze's extension theorem, we can extend g_1 to a function $g_2 \in C(C_1, R^{2n})$ with $|||g_2|||_{C_1} < r/2$. Now set

$$g(x)=\begin{cases} g_2(x) & x\in C_1 \\ f(x) & x\in I^n\backslash C_1 \end{cases}.$$

The function g is continuous since $g_2|\partial(C_1)=f|\partial(C_1)$, and $|||f-g|||=|||f-g|||_{C_1}<r/2+r/2=r$. Hence $g\in U$, and since $g|C_2=\Phi$ it follows that $V=\{h\in U: |||g-h|||<\delta\}$ is an open subset of U that contains no one-to-one maps.

Reference

1. W. Hurewicz and H. Wallman, Dimension Theory. Princeton University Press, Princeton, N.J., 1948.

CLASSROOM NOTES

Edited by Deborah Tepper Haimo and Franklin Tepper Haimo

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St Louis, MO 63121.

MODIFIED CONVERGENCE OF TAYLOR SERIES

Robert D. Small

Department of Mathematics, University of New Brunswick, Fredericton, N. B., Canada E3B 5A3

The domain of convergence of the Taylor series that represents the real-analytic function $f(x)$ at $x=a$,

$$f^T(x)=\sum_{n=0}^{\infty} f^{(n)}(a)(x-a)^n/n!, \tag{1}$$

is determined by the convergence of its sequence of partial sums

$$f_N(x)=\sum_{n=0}^{N} f^{(n)}(a)(x-a)^n/n!. \tag{2}$$

Both (1) and (2) converge within the region $|x-a|<r$ and diverge within the region $|x-a|>r$, where r is the distance from a to the nearest singularity of $f(z)$ in the complex z-plane. We propose that a modified sequence of partial sums can be chosen from (1) that converges over a greater domain than does (2). By means of a rearrangement of terms we obtain an effect that is similar in nature to the phenomenon of overconvergence (see **[1]**, for example) where certain subsequences of the partial sums (2) converge to $f(x)$ outside the circle of convergence.

Consider the modified sequence of partial sums

$$\hat{f}_N(x)=\sum_{n=0}^{N} f^{(n)}(a)\binom{N}{n}N^{-n}(x-a)^n, \tag{3}$$

Since $\lim_{N\to\infty}\binom{N}{n}N^{-n}=1/n!$, each term of (3) converges to the corresponding term of (1) as $N\to\infty$. Then, since (1) converges absolutely in the range $|x-a|<r$, (3) converges there as well.

The finite series (3) are derived by means of linear extrapolation over a successively refined mesh. The interval $a\leqslant t\leqslant x$ is partitioned into N equal parts and the approximation to the function $f(t)$ is continued across each subinterval $c\leqslant t\leqslant c^+(x-a)/N$ by the linear extrapolation

$$f\left(c+\frac{x-a}{N}\right) \cong \left(1+\frac{x-a}{N}\frac{d}{dt}\right)f(t)|_{t=c}. \tag{4}$$

Performing this extrapolation N times gives

$$f(x) \cong \left(1+\frac{x-a}{N}\frac{d}{dt}\right)^N f(t)|_{t=a}, \tag{5}$$

which is (3). A familiar example of this series is found by substituting $a = 0$ and $f^{(n)}(a) = 1$, for all n, into (3), which gives the approximation for e^x,

$$\hat{f}_N(x) = \left(1+\frac{x}{N}\right)^N. \tag{6}$$

The improved convergence of (3) over that of (2) may be demonstrated on the function $g(x) = (1+x)^{-1}$. This function has a pole at $x = -1$ so that if $a = 0$, (2) and (3) converge for $|x| < 1$ and (2) diverges for $|x| > 1$. Since $g^{(n)}(0) = (-1)^n n!$, (3) becomes

$$\hat{g}_N(x) = \sum_{n=0}^{N} \frac{(-1)^n N! x^n}{(N-n)! N^n}. \tag{7}$$

For $x \geqslant 1$ we rearrange (7) by setting $N - n = p$ to obtain

$$\hat{g}_N(x) = N!\left(-\frac{x}{N}\right)^N e^{-N/x} - \sum_{p=N+1}^{\infty} \frac{N!}{p!}\left(-\frac{N}{x}\right)^{p-N}. \tag{8}$$

If we use Stirling's formula, $N! \sim \sqrt{2\pi}N^{N+1/2}e^{-N}$, the first term of (8) is asymptotic to

$$(-1)^N\sqrt{2\pi N}\, e^{N(\log x - 1/x - 1)}.$$

This converges to zero for $\log x - 1/x - 1 < 0$ and diverges otherwise. This condition has been determined numerically to be $x < 3.591121477$. The second term may be written

$$\sum_{p=N+1}^{\infty} (-1)^{p-N+1} x^{-(p-N)} \prod_{j=1}^{p-N} \frac{N}{p-j+1}.$$

For fixed N it is a convergent alternating series and hence it is less in magnitude than its first term $N/(N+1)x$. This bound on the series converges to x^{-1} as $N \to \infty$, and hence (8) converges when its first term does. For $x \leqslant -1$ we note that (7) is composed of positive terms so that reordering them does not affect the convergence. Since (2) diverges in this range, (7) must also. In summary, the reordering of the terms of the Taylor series from (2) to (3) has extended the domain of convergence from $|x| < 1$ to $-1 < x < 3.5911$.

It is noted in closing that an extrapolation of any order may be used in place of (4) to produce a modified sequence of partial sums of (1) that presumably has similar convergence properties to (3). If a polynomial of degree M is used to extrapolate the approximation of $f(t)$ from c to $c + (x-a)/N$ we replace (4) by

$$f\left(c+\frac{x-a}{N}\right) \cong \left(1+\frac{x-a}{N}\frac{d}{dt}+\cdots+\frac{1}{M!}\left(\frac{x-a}{N}\right)^M\frac{d^M}{dt^M}\right)f(t)|_{t=c}$$

and hence the final formula may be written

$$\hat{f}_N{}^M(x) = \left[\sum_{m=0}^{M}\frac{1}{m!}\left(\frac{x-a}{N}\frac{d}{dt}\right)^m\right]^N f(t)|_{t=a}. \tag{9}$$

The equivalent form to (3) is

$$\hat{f}_N{}^M(x) = \sum_{n=0}^{MN} f^{(n)}(a) b_n N^{-n}(x-a)^n$$

where the b_n are computed from $(1+s+s^2/2!+\cdots+S^M/M!)^N=\Sigma_{n=0}^{MN}b_ns^n$. As an example, the approximation of e^x is given by

$$\hat{f}_N{}^M(x)=\left[1+\frac{x}{N}+\frac{1}{2!}\left(\frac{x}{N}\right)^2+\cdots+\frac{1}{M!}\left(\frac{x}{N}\right)^M\right]^N.$$

Reference

1. E. Hille, Analytic Function Theory, vol. 2, Ginn, Boston, 1962.

EQUIVALENCE OF AN IDENTITY IN VECTOR ANALYSIS TO QUATERNION ASSOCIATIVITY, AND RAMIFICATIONS

ANDREW SOBCZYK
Department of Mathematical Sciences, Clemson University, Clemson, SC 29631

Real Quaternions. Identifying the unit vectors **i**, **j**, **k** with quaternion imaginaries, each real quaternion **q** has the form $a+\mathbf{A}$, with scalar component a and vector (or imaginary) component **A**. In terms of the usual dot and cross products, the product **AB** of imaginary quaternions is easily seen to be $-\mathbf{A}\cdot\mathbf{B}+\mathbf{A}\times\mathbf{B}$. Thus

$$(a+\mathbf{A})(b+\mathbf{B})=ab+b\mathbf{A}+a\mathbf{B}-\mathbf{A}\cdot\mathbf{B}+\mathbf{A}+\mathbf{B},$$

and the associator

$$\begin{aligned}\langle(a+\mathbf{A})(b+\mathbf{B})\rangle(c+\mathbf{C})-(a+\mathbf{A})\langle(b+\mathbf{B})(c+\mathbf{C})\rangle&=0+\mathbf{0}\\ =(\mathbf{A}\cdot\mathbf{B})\mathbf{C}-(\mathbf{A}\times\mathbf{B})\times\mathbf{C}-(\mathbf{B}\cdot\mathbf{C})\mathbf{A}&+\mathbf{A}\times(\mathbf{B}\times\mathbf{C}),\end{aligned}$$

or

$$(\mathbf{A}\times\mathbf{B})\times\mathbf{C}-\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\cdot\mathbf{B})\mathbf{C}-\mathbf{A}(\mathbf{B}\cdot\mathbf{C}).\qquad(*)$$

See, e.g., [**3**, (33) and (34), p. 290]. Conversely, identity (*) implies that multiplication of quaternions is associative.

It is well known that the algebra of real 3-vectors with the cross product is the Lie algebra of the real three-dimensional orthogonal group;

$$\text{if } \mathbf{i}=\begin{pmatrix}0&0&0\\0&0&-1\\0&1&0\end{pmatrix},\quad \mathbf{j}=\begin{pmatrix}0&0&1\\0&0&0\\-1&0&0\end{pmatrix},\quad \mathbf{k}=\begin{pmatrix}0&-1&0\\1&0&0\\0&0&0\end{pmatrix},$$

then the Lie algebra also may be described as the algebra of 3 by 3 real antisymmetric matrices $A,B,\ldots$, with the commutator or Lie bracket $[\mathbf{A},\mathbf{B}]=AB-BA=\mathbf{A}\times\mathbf{B}$ as product. We have $[\mathbf{i},\mathbf{j}]=\mathbf{i}\times\mathbf{j}=\mathbf{k}$, $\mathbf{j}\times\mathbf{k}=\mathbf{i}$, $\mathbf{k}\times\mathbf{i}=\mathbf{j}$.

Change to $\mathbf{j}\times\mathbf{k}=-\mathbf{i}$. The Lie algebra of the real two-dimensional special linear group $SL_2(R)$ is the algebra of all real 2-by-2 matrices having trace 0, with the commutator product. (See, e.g., [**4**, p. 89].) Choosing as basis

$$\mathbf{i}=\begin{pmatrix}0&1\\-1&0\end{pmatrix},\quad \mathbf{j}=\begin{pmatrix}-1&0\\0&1\end{pmatrix},\quad \mathbf{k}=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$

we have $[\mathbf{i},\mathbf{j}]=\mathbf{k}$, $[\mathbf{j},\mathbf{k}]=-\mathbf{i}$, $[\mathbf{k},\mathbf{i}]=\mathbf{j}$, the same as the cross-product algebra except that $[\mathbf{j},\mathbf{k}]=-\mathbf{i}$ instead of $[\mathbf{j},\mathbf{k}]=+\mathbf{i}$. The effect of the change of sign is to make the analogous algebra to the quaternion algebra not associative: $(\mathbf{ij})\mathbf{k}=\mathbf{k}^2=-1$ but $\mathbf{i}(\mathbf{jk})=-\mathbf{i}^2=+1$. Like the quaternions, this latter algebra is noncommutative; also there are non-null divisors of zero, e.g., $(1+\mathbf{i}+\mathbf{j}+\mathbf{k})\,(1+\mathbf{i}-\mathbf{j}+\mathbf{k})=0$ but $(1+\mathbf{i}-\mathbf{j}+\mathbf{k})\,(1+\mathbf{i}+\mathbf{j}+\mathbf{k})=4(\mathbf{i}+\mathbf{k})$. See [**1**], [**5**].

Vectors with Complex Components. In order to consider complex 3-vectors $\mathbf{A}_1+i\mathbf{A}_2$,

$\mathbf{B}_1 + i\mathbf{B}_2,\ldots$, where $\mathbf{A}_i$, $\mathbf{B}_i$, etc., are real vectors, to avoid confusion we now change notation from **i, j, k** to, respectively, e_1, e_2, e_3, freeing i to be $\sqrt{-1}$. For $\mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2$, $\mathbf{B} = \mathbf{B}_1 + i\mathbf{B}_2$, the product

$$\mathbf{A}\cdot\mathbf{B} = (\mathbf{A}_1 + i\mathbf{A}_2)\cdot(\mathbf{B}_1 + i\mathbf{B}_2) = (\mathbf{A}_1\cdot\mathbf{B}_1 - \mathbf{A}_2\cdot\mathbf{B}_2) + i(\mathbf{A}_2\cdot\mathbf{B}_1 + \mathbf{A}_1\cdot\mathbf{B}_2)$$

is the negative of the scalar part, and $\mathbf{A}\times\mathbf{B} = (\mathbf{A}_1 + i\mathbf{A}_2)\times(\mathbf{B}_1 + i\mathbf{B}_2)$ is the quaternion imaginary part, of the product of imaginary complex quaternions

$$(A_{11} + iA_{21})e_1 + (A_{12} + iA_{22})e_2 + (A_{13} + iA_{23})e_3$$

and

$$(B_{11} + iB_{21})e_1 + (B_{12} + iB_{22})e_2 + (B_{13} + iB_{23})e_3.$$

As in the case of real quaternions, associativity of multiplication of complex quaternions is equivalent to identity (*) in which **A**, **B**, **C** are allowed to be complex 3-vectors. (The complex quaternions are an associative algebra, of which the underlying real linear space is eight-dimensional, and which has divisors of zero, e.g., $(i + e_1)(i - e_1) = i^2 - e_1^2 = -1 + 1 = 0$.) In case of complex 3-vectors with complex orthogonality, there are isotropic vectors, i.e., non-null vectors that are self-orthogonal, e.g., $\mathbf{A}\cdot\mathbf{A} = |\mathbf{A}_1|^2 - |\mathbf{A}_2|^2 + 2i\mathbf{A}_1\cdot\mathbf{A}_2 = 0$ if $|\mathbf{A}_1| = |\mathbf{A}_2|$ with $\mathbf{A}_1\cdot\mathbf{A}_2 = 0$. To make $\mathbf{A}\times\mathbf{B}$ Hermitian-orthogonal to **A** and to **B**, we may define the cross product by $\mathbf{A}\otimes\mathbf{B} = \bar{\mathbf{A}}\times\bar{\mathbf{B}}$, where $\bar{\mathbf{A}}$ denotes the complex conjugate vector $\mathbf{A}_1 - i\mathbf{A}_2$ of $\mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2$, and $\times$ is the usual real cross product. The Hermitian inner product is left invariant by unitary transformations. The Lie algebra of the unitary group is the commutator algebra of the anti-Hermitian 3-by-3 matrices, which is not isomorphic with the algebra of complex antisymmetric matrices with $\otimes$ as product; the latter algebra is merely the image under conjugation of the complexification of the algebra of the three-dimensional orthogonal group. If a complex matrix $A = A_1 + iA_2$ is anti-Hermitian, then A_1 must be antisymmetric and A_2 symmetric. Since the product of a real antisymmetric matrix by a complex number is not anti-Hermitian, it is clear that the Lie algebra of the unitary group is only a real Lie algebra; i.e., it is not closed under scalar multiplication by complex scalars. Apparently if there is an extension to complex vectors of the relation between real quaternions and the cross product of real vectors, having the desired property of Hermitian orthogonality, it is not a trivial extension.

Cayley Numbers. Cayley numbers or octonions are a nonassociative real eight-dimensional algebra, without zero divisors, which has as a basis the scalar unit 1 and seven imaginaries $\{e_i\}$, $i = 1,\ldots,7$. For a description, see, e.g., [2]. Each of the triples $\{e_1,e_2,e_4\}$, $\{e_2,e_3,e_5\}$, $\{e_3,e_4,e_6\}$, $\{e_4,e_5,e_7\}$, $\{e_5,e_6,e_1\}$, $\{e_6,e_7,e_2\}$, $\{e_7,e_1,e_3\}$ is analogous to $\{i,j,k\}$ and, with 1, forms a basis for a real quaternionic subalgebra. The triples are a "Steiner system" that covers all of the pairs $\{e_i,e_j\}$, each exactly once. The multiplication table is $e_i^2 = -1$, $i = 1,\ldots,7$, with the product of the three imaginaries in each triple also -1, e.g., $e_1e_2e_4 = -1$, which is shorthand for $e_2e_4 = -e_4e_2 = e_1, e_4e_1 = -e_1e_4 = e_2, e_1e_2 = -e_2e_1 = e_4$. The seven triples are associative; all other triples are not, e.g., $e_1(e_2e_5) = e_1(-e_3) = -e_7$, but $(e_1e_2)e_5 = e_4e_5 = +e_7$. The algebra L_7 of seven-dimensional vectors $A = \Sigma_{i=1}^{7} a_ie_i = \{A_1,A_2,\ldots,A_7\}$, where $A_1,A_2,\ldots,A_7$ are 3-vectors $A_1 = (a_1,a_2,a_4)$, $A_2 = (a_2,a_3,a_5)$, etc., with product defined by $e_i^2 = 0$,

$$A\times B = A_1\times B_1 + A_2\times B_2 + \cdots + A_7\times B_7,$$

is a Lie algebra that bears a similar relationship to the algebra of Cayley numbers, to that of the real 3-vectors with $\times$ to the quaternions; i.e., the Cayley numbers may be regarded as $a + A$, $b + B$, etc., where a, b, etc., are the scalar, and A, B, etc., the seven-dimensional, purely imaginary components of the octonions.

References

1. S. C. Althoen and J. F. Weidner, Real division algebras and Dickson's construction, this MONTHLY, 85 (1978) 368–371.

2. H. S. M. Coxeter, Integral Cayley numbers, Duke Math. Journal, 13 (1946) 561–578.
3. F. B. Hildebrand, Advanced Calculus for Engineers, Prentice-Hall, Englewood Cliffs, N.J., 1956.
4. Serge Lang, $SL_2(R)$, Addison-Wesley, Reading, Mass., 1975.
5. R. D. Schafer, An Introduction to Nonassociative Algebras, Academic Press, New York, 1966.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

Send all ***proposed*** *problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Product Divided by LCM

S 21 [1979, 784]. *Proposed by Paul Erdős, Technion, Haifa, Israel.*

Let $A(n,k)=(n+1)(n+2)\cdots(n+k)$, $B(n,k)=\operatorname{lcm}\ [n+1,n+2,\ldots,n+k]$, and $\alpha(n,k)=A(n,k)/B(n,k)$.

(a) How many distinct values can $\alpha(n,k)$ take for fixed k?

(b)* Do m, n, k exist with $m>n+k-1$ and $B(m,k)=B(n,k)$?

Solution to (a) by Albert Nijenhuis, University of Pennsylvania. Since $\alpha(n,1)=1$ for all n, we assume that $k>1$. Let p be any prime and $m_p(x)$ the multiplicity of p in x: $p^{m_p(x)}|x$ and $p^{1+m_p(x)}\nmid x$; let m_p be maximal at some x_0 in $[n+1,n+k]$, and then $m_p(B(n,k))=m_p(x_0)$ and $m_p(x)=m_p(x-x_0)$ for all $x\in[n+1,n+k]-\{x_0\}$; so $m_p(\alpha(n,k))=m_p((n+k-x_0)!(x_0-n-1)!)$. Let $\beta_p(k)$ be the number of values of $m_p(k'!(k-k'-1)!)$ for $k'\in[0,k-1]$; then the number of distinct values of $\alpha(n,k)$ equals $\Pi\{\beta_p(k)|\,p \text{ prime}, p<k\}$. Now it is well known that $m_p(x!y!)=m_p((x+y)!)-c_p(x,y)$ (Kummer, 1852), where $c_p(x,y)=m_p(\binom{x+y}{x})$ is the number of "carries" in the addition $(x,y)\to x+y$ when performed in base p. The number of digits of $k-1$ to base p is $1+[\log_p(k-1)]$; let the least significant s of these ($s\geqslant 0$) be equal to $p-1$, and then carries in additions with sum $k-1$ can take place in any of the remaining digits except of course the most significant one. Since $s=m_p(k)$, we see that the number of different values of $c_p(k',k-k'-1)$ is therefore $1+[\log_p(k-1)]-m_p(k)$. An exception occurs when all digits of $k-1$ equal $p-1$; i.e., where k is a power of p. It is easy to see, however, that $\beta_p(k)=1+[\log_p(k)]-m_p(k)$ in all cases. Using the Chinese Remainder Theorem, one sees that the number of distinct values of $\alpha(n,k)$, for fixed k, is the product $\Pi([\log_p k]-m_p(k)+1)$ of the $\beta_p(k)$ for all primes $p<k$.

Part (a) was also solved by A. Sakmar (France) and the proposer.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by October 31, 1981. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2890. *Proposed by Barry J. Powell, Kirkland, Washington.*

Let $p > 3$ be a prime for which $2p - 1 = q$ is also prime. Let $B_{2m} = N_{2m}/D_{2m}$ represent the $2m$th Bernoulli number, where N_{2m} is the numerator and D_{2m} is the denominator (in usual notation). Prove that $|N_{2p-2}| > |N_{2p}|$.

*This suggests that there are infinitely many positive integers m for which $|N_{2m-2}| > |N_{2m}|$. Can this be proved?

E 2891. *Proposed by Abou-Dardaye Barry, Gabon, Central Africa.*

In "250 problems in elementary number theory," problems 140, 141, Sierpiński shows that the Diophantine equation $x(x+1) = k^2y(y+1)$ has no solution with $xy \neq 0$ if $k = 2$ or if $k = p^n$. Find other values of k that give the same conclusion.

E 2892. *Proposed by F. S. Cater, Portland State University.*

Let X be a nonvoid set; let $k > 0$ be an integer. Characterize the families U of subsets (of X), card $U = 2^k$, satisfying (*) for every $A \in U$ there are exactly k sets $B \in U$ such that $(A \setminus B) \cup (B \setminus A)$ is a singleton. (Compare E 2792 [1979, 702].)

E 2893. *Proposed by Ko-Wei Lih, Institute of Mathematics, Academia Sinica, Taipei.*

Determine all homeomorphisms f from $[0,1]$ to $[0,1]$ that are solutions of the functional equation $f(2x - f(x)) = x$ for all x in $[0,1]$.

E 2894*. *Proposed by Thomas Ihringer, Technische Hochschule, Darmstadt, Germany.*

Let n be fixed. In how many ways can a square be dissected into (a) n congruent rectangles, (b) n rectangles of equal area?

E 2895. *Proposed by M. Vulis, CUNY Graduate Center.*

The children at a birthday party sit around a table and divide a homogeneous cake according to the following rules: Initially the birthday child divides the cake into two equal portions, giving one portion to one of the two neighboring children and keeping the other. At the next and succeeding steps, some pair of adjacent children combine their portions and divide equally. (Thus if two adjacent children have portions a and b before division and share them, each has a portion $\frac{1}{2}(a+b)$ after the division. Note that either a or b can be zero.)

The process terminates in a finite number of steps with a fair division of the cake. How many children are at the party?

SOLUTIONS OF ELEMENTARY PROBLEMS

The sum $\Sigma(k,n)$

E 2821 [1980, 220]. *Proposed by Jeffrey Shallit, undergraduate, Princeton University.*

Express $S(n) = \Sigma_{k=1}^{n} \gcd(k,n)$ in terms of the factorization of n as a product of powers of distinct primes.

Solution. $S(k,n) = \Sigma_{d|n} d\phi(n/d)$. Since S is a Dirichlet convolution, S is multiplicative and if $n = \prod P^{\alpha}$, $S(n) = \prod P^{\alpha-1}(\alpha P + P - \alpha)$. ($P$ is the typical prime dividing n; α is its exponent.)

F. Eugenio (Italy), S. S. Wang, and R. Sivarama Krishnan (India) mentioned $\Sigma(k,n)^{\lambda}$. The latter referred to S. S. Pillai, *On an arithmetical function*, J. Annamalai Univ., 2, no. 2 (1933) 243–248. O. G. Ruehr and J. Schwaiger (Austria) referred to I. Niven and H. Zuckerman, *Number Theory* (Wiley), p. 96, no. 6. F. G. Schmitt, Jr., referred to Andrews's *Number Theory*, p. 91, problem 10.

The solvers were H. L. Abbott (Canada), M. Bates, R. Beigel, K. A. Beres, R. Breusch, P. S. Bruckman, F. S. Cater, M. P. Eisner, L. L. Foster, N. Franceschine III, A. L. Furno, R. Gilmer, N. Glick, E. Grosswald, D. Hensley, R. T. Hood, M. Josephy (Costa Rica), W. T. M. Kars (Holland), J. B. Klerlein & A. G. Starling, K. A. Klinger, M. F. Kruelle (student), L. Kuipers (Switzerland), B. Kurtzman, O. P. Lossers (Netherlands), D. L. Mabbott, G. Mason & G. Viswanathan (Canada), L. Matejcka (Czechoslovakia), M. McConnell, J. L. de Miguel (Spain), N. Miku (Netherlands), W. Myers, R. Patenaude, B. Peterson, B. Prielipp, H. Prodinger (Austria), P. Schumer, S. Seltzer, R. W. Sheets, A. Smuckler (Israel), L. Somer, J. Suck (Germany), D. Thoro, E. Trost (Switzerland), L. van Hamme (Belgium), D. Weisser, M. Woltermann, K. L. Yocum, and the proposer.

The Equation $x^{2p} + y^{2p} + z^{2p} = w^{2p}$

E 2824 [1980, 220]. *Proposed by Barry J. Powell, Kirkland, Wash.*

Prove that for p any odd prime with $p \not\equiv 1 \pmod 4$, the equation

$$x^{2p} + y^{2p} + z^{2p} = w^{2p} \tag{*}$$

has no solution in positive integers x,y,z,w with $xyzw \not\equiv 0 \pmod p$.

(Compare this problem with E 2771 [1979, 308].)

Solution by Robert Breusch, Amherst, Mass., and L. L. Foster, California State College, Northridge (independently). Assume, per contra, that $(x,y,z,w) = 1$ and (*) holds. Reducing modulo 8, we may assume that x,y are even, z,w are odd. From (*),

$$x^{2p} + y^{2p} = (w^2 - z^2)K,$$

$$K = \sum_{i=0}^{p-1} w^{2(p-1-i)} z^{2i} = pw^{2(p-1)} + w^{2(p-2)}(z^2 - w^2) + w^{2(p-4)}(z^4 - w^4) + \cdots. \tag{**}$$

Since Σ has p terms, $K \equiv 3 \pmod 4$. Therefore, K contains at least one prime factor q, $q \equiv 3 \pmod 4$, with odd exponent. From [**1**, p. 117, Ex. 6], $q|x$ and $q|y$. Thus $q \neq p$, and $x^{2p} + y^{2p}$ contains q with even exponent. Thus from (**), $q|w^2 - z^2$. Thus, from the second formula for K, $q|pw^{2(p-1)}$. Therefore $q|pw, q|w$, and finally (since $q|w^2 - z^2$), $q|z$. This descent contradicts $(x,y,z,w) = 1$.

Reference

1. I. Niven & H. Zuckerman, *Theory of Numbers*, 3rd ed., Wiley.

Also solved by the proposer.

Abel Convolution Identity

E 2828 [1980, 303]. *Proposed by Jerrold W. Grossman and Hai-Ping Ko, Oakland University, Rochester, Mich.*

Prove that $\Sigma C_i^n (i+1)^{i-1}(j+1)^{j-1} = 2(n+2)^{n-1}$, the sum being extended over all nonnegative integers i,j such that $i + j = n$.

Solution. The book *Combinatorial Identities* by H. W. Gould lists the "generalized Abel convolution"

$$(x+y)\sum_{k=0}^{n} B_k(x,z)B_{n-k}(y,z)(r+sk) = [r(x+y)+snx]B_n(x+y,z),$$

where B is defined by

$$k!B_k(x,z) = x(x+kz)^{k-1}.$$

Set $s=0$, $r=z=1$; $x=y=1$. □

Comments. L. Kuipers (Switzerland) used this formula directly. O. G. Ruehr (Michigan Technological University) derived a generalization; he also referred to Gould. W. A. Al-Salam (University of Alberta, Canada) and I. G. Macdonald (Queen Mary College, London, U.K.) used a Lagrange expansion formula and mentioned the generalization

$$\alpha_\beta\cdots\zeta\sum\binom{n}{i,j,\ldots,p}(\alpha+i)^{i-1}(\beta+j)^{j-1}\cdots(\zeta+p)^{p-1} = k(n+k)^{n-1};$$

the multiple sum is on $i,j,\ldots,p$, subject to $i+j+\cdots+p=n$, and $\alpha+\beta+\cdots+\zeta=k$.

M. D. Ašić (student, Yugoslavia) and M. Wolterman counted trees. O. P. Lossers (Netherlands) gave two proofs; he and M. Vowe (Switzerland) referred to J. Riordan's *Combinatorial Identities*. K. L. Bernstein referred to this MONTHLY, 1978, p. 452.

Also solved by P. Bracken (Canada), I. Gessel, V. Hernandez (Spain), K. A. Klinger, and the proposers.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by October 31, 1981. The solver's full post-office address should be on each sheet.

6347. *Proposed by David E. Daykin, University of Reading, England, and Roland Häggkvist, Institut Mittag-Leffler, Sweden.*

Prove or disprove that any partial Latin square can be partitioned into two partial Latin squares both of which can be independently completed. For example,

123x
xxx4
xxxx
xxxx

cannot be completed, but each of

123x,	xxxx
xxxx	xxx4
xxxx	xxxx
xxxx	xxxx

can be completed.

6348. *Proposed by James Henle and Stanley Wagon, Smith College.*

Prove or disprove: If μ is a countably additive translation invariant measure from the Borel subsets of $\mathbb{R}$ to $[0,\infty]$, then μ is invariant under all isometries of $\mathbb{R}$; i.e., it is also invariant under reflections.

6349. *Proposed by J.-B. Hiriart-Urruty, University of Clermont-Ferrand II, France, and T. H. Sweetser, University of California at San Diego.*

Let $\mathcal{Q}$ be a countable set in $L(\mathbb{R}^n,\mathbb{R}^m)$—the space of all linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$; we define $\mathcal{Q}x = \{Ax \mid A \in \mathcal{Q}\}$. Does $B \in L(\mathbb{R}^n,\mathbb{R}^m)$ satisfying

$$Bx \in \mathcal{Q}x \quad \forall x \in \mathbb{R}^n$$

necessarily belong to $\mathcal{Q}$?

SOLUTIONS OF ADVANCED PROBLEMS

Injective Lie Algebras

6169 [1977, 659]. *Proposed by Joseph Rotman, University of Illinois.*

Prove that the category of all Lie algebras over a field K has no injective objects other than 0.

Solution by J. Humphreys, University of Massachusetts. If $L \neq 0$ were an injective object in the category of Lie algebras over a field K, then L would have to be an ideal direct summand in any larger Lie algebra. It is well known that L acts faithfully on some vector space V (cf. N. Jacobson, *Lie Algebras*, Interscience, New York, 1962, p. 162). Form the semidirect sum $L' = L + V$ (V an abelian ideal). Since $L \neq 0$, $0 \neq [LV] \subset V$, so L cannot be an ideal in L', contrary to assumption.

Also solved by A. A. Jagers (Netherlands) and the proposer.

Complementary Subsets of the Irrationals

6188 [1978, 53]. *Proposed by F. S. Cater, Portland State University.*

Do there exist complementary subsets A and B of the set of irrational numbers such that for any open intervals I and J in the real line, (1) $A \cap I$ and $B \cap J$ are not homeomorphic in the Euclidean topology, and (2) there is a one-to-one continuous function mapping $A \cap I$ onto $B \cap J$?

Solution by the proposer. The answer to the question is "yes."

In what follows, Q = set of rational numbers, X = set of irrational numbers, $R = X \cup Q$ = the real line.

LEMMA 1. *Let W be any countable subset of X. Then $X \setminus W$ is homeomorphic to X.*

Proof. Q and $Q \cup W$ are both countable sets dense in R. It is easy to construct an increasing mapping f of Q onto $Q \cup W$. Then f can be extended in a natural manner to a continuous increasing function g mapping R onto R. It follows that g is an increasing homeomorphism of R onto R, and $g(X) = g(R \setminus Q) = R \setminus g(Q) = X \setminus W$.

LEMMA 2. *Let J be an open interval of R. Then there do not exist disjoint dense subsets U and V of J such that U is a G_δ-set and V is homeomorphic to X.*

Proof. Suppose such U and V exist. Then there exist open sets W_n such that $\bigcap_n W_n = U$. But each $W_n \cap V$ is an open dense subset of the space V and $\bigcap_n (W_n \cap V) = (\bigcap_n W_n) \cap V = U \cap V = \varnothing$. This conflicts with the fact that X is of second category.

Now let C be a dense G_δ-set in R with measure 0. Let (I_n) denote a sequence listing all open intervals in R with endpoints in Q. By a theorem of Mazurkiewicz (see Thm. 85 of W. Sierpiński: *Introduction to General Topology*, 1934), for each n there is a countable subset C_n of $C \cap I_n$ such that $(C \setminus C_n) \cap I_n$ is homeomorphic to X. Put $U = C \setminus [Q \cup (\bigcup_m C_m)]$. Then by Lemma 1, $U \cap I_n$ is homeomorphic to X for each n.

Suppose (a, b) is any open interval of R. Choose rational numbers a_n such that $< \cdots < a_{-n} < a_{1-n} < \cdots < a_{-1} < a_0 < a_1 < a_2 < \cdots < a_n < \cdots$ and $a_{-n} \downarrow a$, $a_n \uparrow b$. Then each $U \cap (a_{n-1}, a_n)$ is homeomorphic to X, and it follows that $U \cap (a, b) = \bigcup_{n=\infty}^{\infty} U \cap (a_{n-1}, a_n)$ is homeomorphic to X.

Put $D = R \setminus C$. By a known result (W. Sierpiński, ibid., Thm. 86), for each n there is a countable subset D_n of $D \cap I_n$ such that there is a continuous one-to-one mapping of X onto

$(D \setminus D_n) \cap I_n$. Put $V = D \setminus [Q \cup (\bigcup_m D_m)]$. Then by Lemma 1, there is a one-to-one continuous mapping of X onto $V \cap I_n$ for each n.

Suppose (a, b) is any open interval of R. Choose rational numbers a_n as before. Then there is a continuous one-to-one mapping of X onto each $V \cap (a_{n-1}, a_n)$, and consequently there is a one-to-one continuous mapping of X onto $V(a, b) = \bigcup_{n=\infty}^{\infty} V \cap (a_{n-1}, a_n)$.

Thus for any open intervals I and J of R, there is a homeomorphism mapping $U \cap I$ onto X and a continuous one-to-one mapping of X onto $V \cap J$. So there is a continuous one-to-one mapping of $U \cap I$ onto $V \cap H$. But $V \cap J$ cannot be homeomorphic to $U \cap I$; for if it were, $V \cap J$ would be homeomorphic to X, and the sets $V \cap J$ and $U \cap J$ would conflict with Lemma 2.

Now $U \cap V = \varnothing$ and $R \setminus (U \cup V)$ is evidently a countable set containing Q. There is an increasing mapping f of $R \setminus (U \cup V)$ onto Q. We extend f to an increasing homeomorphism g of R onto R, and put $A = gU$, $B = gV$. Then $A \cap B = \varnothing$ and $A \cup B = R \setminus Q = X$. In other words, A and B are complementary subsets of X. Finally, A and B inherit properties (1) and (2) from the sets U and V.

Also solved by M. E. Rudin.

Invertible Laurent Polynomials

6259 [1979, 226]. *Proposed by William D. Blair, Northern Illinois University, and James E. Kettner, Hanover Park, Ill.*

Let R be a commutative ring with unity and let $R[x, x^{-1}]$ be the ring of Laurent polynomials $f(x) = \Sigma_{i=-m}^{n} a_i x^i$ over R. Find necessary and sufficient conditions on the coefficients a_i of $f(x)$ for $f(x)$ to be invertible.

Solution by the proposers. First we observe that if R is an integral domain then a degree argument similar to the one used for ordinary polynomials shows that $f(x)$ is invertible if and only if all coefficients a_i are zero except one, which is invertible. In general we show that $f(x)$ is invertible if and only if $\Sigma_{i=n}^{m} a_i^2$ is a unit of R and the products $a_i a_j$ are all nilpotent for distinct i and j. If $f(x)$ is invertible then $\bar{f}(x) = \Sigma_{i=n}^{m} \bar{a}_i x^i$ is invertible in $R/P[x, x^{-1}]$ for all prime ideals P of R, and so for each prime ideal P, $a_i a_j \in P$ if $i \neq j$ since at least one of a_i or a_j belongs to P, and thus $a_i a_j$ is nilpotent. Also, if Σa_i^2 were not a unit then there would exist a prime ideal P such that $\Sigma a_i^2 \in P$. But since all but one a_i, say a_{i_0}, belongs to P and a_{i_0} is invertible modulo P we have a contradiction. Conversely, if Σa_i^2 is a unit of R and $a_i a_j$ is nilpotent for distinct i and j, then we consider the product $(\Sigma_{i=n}^{m} a_i x^i)(\Sigma_{i=n}^{m} a_i x^{-i})$. The coefficients of x^i, $i \neq 0$ are all nilpotent and the coefficient of x^0 is a unit of R. Thus the product is a unit of R plus a nilpotent element of $R[x, x^{-1}]$ and so $f(x)$ is invertible in $R[x, x^{-1}]$.

Also solved by Robert Gilmer, Bruce Glastad, Charles C. Hanna, A. A. Jagers (Netherlands), David Lantz, James R. Smith, and Edward T. Wong.

Several solvers give variants of the solution above. For example, we may replace the condition "Σa_i^2 is a unit" by "Σa_i is a unit."

Gilmer, Glastad, Hanna, Jagers, and Wong give conditions similar to the one above. Gilmer and Glastad refer to the paper *The group of units of a commutative semigroup ring*, by Gilmer and Raymond C. Heitmann, to appear in Pacific J. Math.

Lantz and Smith give nearly the same characterization, that, for every prime ideal P, all but precisely one a_i are in P. See Lantz, *R-automorphisms of* $R[G]$ *for G abelian torsion-free*, Proc. Amer. Math. Soc., 61 (1976) 1–6.

Conditions for Unique Factorization

6264 [1979, 310]. *Proposed by William C. Waterhouse, The Pennsylvania State University.*

Despite the commonly repeated story, it appears that Kummer never actually made the mistake of presuming unique factorization in rings of cyclotomic integers. In a paper withdrawn before publication, however, he did once assume that when he had two elements with no common factor

he could write 1 as a linear combination of them. [See H. M. Edwards, Arch. Hist. Exact Sci. 14 (1975) 219–236, and 17 (1977) 381–394.] Show that for a Noetherian integral domain this assumption implies unique factorization.

Solution by Robert Gilmer, Florida State University. Let D be a Noetherian integral domain with the given property, and let P be a nonzero proper prime ideal of D. Choose a nonzero element x of P. Since D is Noetherian, x is a finite product of irreducible elements, and hence P contains an irreducible element d. If $y \in D - (d)$, then d and y have no common nonunit factor, so $D = (y, d)$. Consequently, (d) is maximal in D, so $P = (d)$ is principal. Therefore each prime ideal of D is principal, and it is known that this implies that D is a principal ideal domain, hence a unique factorization domain.

Essentially equivalent solutions were given by Daniel D. Anderson, F. S. Cater, P. M. Cohn (England), Thomas C. Craven, R. O. Davies (England), Hugh M. Edgar, Michael Fraser, Lorraine L. Foster, R. M. Guralnick, Jay Hook, Miguel L. Laplaza (Puerto Rico), R. Sivaramakrishnan (India), Earl J. Taft, John J. Watkins, and the proposer.

Commuting Automorphisms

6277 [1979, 709]. *Proposed by Yasuo Watatani, Osaka Kyoiku University.*

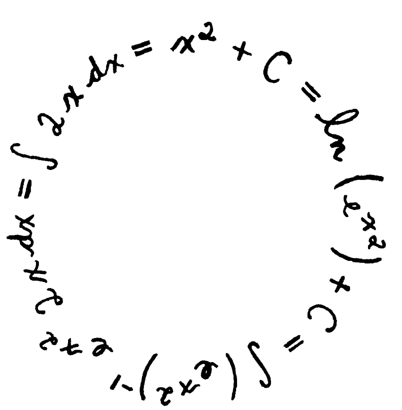
If α and β are $*$-automorphisms of the algebra $B(H)$ of all bounded linear operators acting on a Hilbert space H such that $\alpha(x) + \alpha^{-1}(x) = \beta(x) + \beta^{-1}(x)$ for $x \in B(H)$, then they commute.

Solution by the proposer. Since every $*$-automorphism of $B(H)$ is inner, there are unitaries u and v such that $\alpha(x) = uxu^*$ and $\beta(x) = vxv^*$ for every $x \in B(H)$. By the hypothesis we have $uxu^* + u^*xu = vxv^* + v^*xv$. Putting $x = u$ we have $2u = vuv^* + v^*uv$, that is, u is an average of two unitaries. However the well-known theorem of R. Kadison tells us that every unitary is an extreme point of the unit ball of $B(H)$, so that we have $uv = vu$. Hence α and β commute.

---

57. MISCELLANEA

RUNAROUND



—Ray Bobo, Department of Mathematics, Georgetown University

REVIEWS

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

We invite readers to submit reviews of significant recent college-level mathematics books. We especially encourage reviews based on classroom use and comparative reviews of several related books. Reviews should ordinarliy not exceed two pages (per book) typed double spaced. Manuscripts of reviews as well as books submitted for review should be sent to: Book Review Editor, American Mathematical Monthly, St. Olaf College, Northfield, MN 55057.

The William Lowell Putnam Mathematical Competition, Problems and Solutions: 1938–1964. By A. M. Gleason, R. E. Greenwood, and L. M. Kelly. Mathematical Association of America, 1980. xi + 652 pp.

Which of these two problems do you like better, (1) or (2)? (1) Show that the gravitational attraction exerted by a thin homogeneous spherical shell at an external point is the same as if the material of the shell were concentrated at the center. (2) Let S be a set of $n > 0$ elements, and let $A_1, A_2, \ldots, A_k$ be a family of distinct subsets, with the property that any two of these subsets meet. Assume that no other subset of S meets all of the A_i. Prove that $k = 2^{n-1}$.

If your answer is (1), you are more nearly in tune with the Putnam examination committee for 1938 than with the one for 1964. Each one of the first five examinations has a question about mechanics (1938–1942), but none of the last five does (1960–1964). The word "set" occurs five times in the twenty-fifth examination, and so does "sequence"; neither word occurs in any of the first five examinations. The vocabulary of the first five examinations includes expressions such as "orthogonal trajectories of a family of conics" and "families of rulings on a hyperboloid," which are not found in the last five; the last five mention integral domains and fields, closed sets and dense sets, but the first five never do.

The Putnam competitions were started in 1938 and, except for a three-year gap during World War II, have continued ever since. The entrants are teams and individuals from American and Canadian colleges; the winners get praise, fame, and cash; one of the five top competitors gets a graduate fellowship at Harvard.

The bulk of the volume under review consists of the problems in the first 25 examinations (347 of them) and their solutions (covering well over 500 pages). An appendix contains (among other things) a severe criticism of the examinations by L. J. Mordell, with a rejoinder by L. M. Kelly. Mordell says that the syllabus of the examinations is bad, that the examinations do not stick to the syllabus, that the problems are often too easy or too hard, and that the solutions are often "synthetic" and involve cumbersome arithmetical calculations. Speaking of the problems, he uses words such as pathological, bizarre, undesirable, and unreasonable. Kelly's defense is, in part, a demurrer in the legal sense of the word. (A common colloquial formulation of a demurrer is "So what?" In other words: "While some of what you say may be true, it doesn't prove what you say it does.") Perhaps more to the point, however, is the following comment by Ivan Niven (quoted by Kelly): "It seems to me that Mordell, in criticizing the selection of problems used in Putnam examinations, is really criticizing the current interests, fashions, outlooks, and standpoints of contemporary American mathematics. The difficulty is that what is respectable and normal to one is exotic to another."

I too have a criticism of the Putnam problems, but it is much milder than Mordell's; it is on what might be called the editorial level. The problems, I think, are not always well expressed, and, as a result, they are either stylistically imprecise or, at the very least, just not very attractive. (I could give several examples; problem (2) in the first paragraph could be one of them.) A part of my complaint is against the preference that many examiners have for the imperative mood, as

opposed to the interrogative one. Students are, I think, much more interested in being asked "What is ...?" than in being told to "Prove that ...!" The reaction "Why should I?" to the latter is frequently a perfectly appropriate value judgment.

Here is an example. "Prove that for every real or complex x,

$$\prod_{k=1}^{\infty} \frac{1+2\cos\frac{2x}{3^k}}{3} = \frac{\sin x}{x}."$$

Why should I? Am I alone in thinking that "What are all solutions of $n^m = m^n$ in integers n and m?" is more inviting than "Find all solutions ..."? And what about "Can the unit disk in the plane be partitioned into two disjoint congruent subsets?"—isn't that a lot more interesting and challenging than "Show that it cannot be done"? (The last example, to be sure, is harder as a question than as an order; but doesn't the extra price buy anything of value?)

The examiners sometimes demonstrate that a little innocent merriment doesn't hurt. The fifth competition contains the following two problems. From the morning session: "A square of side $2a$, lying always in the first quadrant of the XY plane, moves so that two consecutive vertices are always on the X- and Y-axes respectively. Find the locus of the midpoint of the square." From the afternoon session: "A square of side $2a$, lying always in the first quadrant of the XY-plane, moves so that two consecutive vertices are always on the X- and Y-axes respectively. Prove that a point within or on the boundary of the square will in general describe a (portion of a) conic. For what points of the square does this locus degenerate?" The solution ends with the following comment. "This problem, the first problem of the afternoon session, is a nice generalization of the first problem of the morning session."

Another piece of mathematical merriment occurs in the twelfth competition. "A mathematical moron is given two sides and the included angle of a triangle and attempts to use the Law of Cosines: $a^2 = b^2 + c^2 - 2bc\cos A$, to find the third side a. He uses logarithms as follows. He finds log b and doubles it; adds to that the double of log c; subtracts the sum of the logarithms of 2, b, c, and $\cos A$; divides the result by 2; and takes the antilogarithm. Although his method may be open to suspicion, his computation is accurate. What are the necessary and sufficient conditions on the triangle that this method should yield the correct result?" (Note the classically correct use of the English neuter pronoun referring to a generic member of the human race.)

Here is a tidbit that most people do not know; it has to do with the Anning-Erdős theorem, which appeared as one of the problems on the eighteenth examination. The theorem says that, if all the distances between pairs of points of an infinite set in the plane are integers, then the set is a subset of a straight line. The Anning-Erdős proof appeared in 1945; it is complicated. When the review copy of the paper reached *Mathematical Reviews*, the then Executive Editor (Ralph Boas) asked Irving Kaplansky (who was, by the way, the first Putnam Scholar at Harvard) to review it. Kaplansky discovered the now known trivial proof. At that time Erdős was in parts unknown and couldn't be reached, so Kaplansky published his proof under Erdős's name, on the theory that Erdős *ought* to have found it. Boas held the review till the new paper came out, reviewed it himself, and signed Kaplansky's name, just to symmetrize things. (I quote this story, almost verbatim, from a letter by Ralph Boas; a pertinent reference is MR7, p. 164.)

The hardest work in producing a book such as this is to look up the solutions, in most cases to think them up, and in all cases to write them up. The authors did all that, and they did it very well. Some of the Putnam problems had mistakes in them, misprints or worse. Some of those are corrected in the present volume, but, such are the gremlins that harass all authors and editors, some new ones sneaked in. Small matter: this is a good book. All problem solvers (that does not quite include all mathematicians, but pretty nearly) will enjoy using it, and some will perhaps enjoy finding fault with it; in the name of all of us, I thank the authors for producing it.

P. R. HALMOS, Indiana University

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, S, L. Recreations in Logic. D.G. Wells. Dover, 1979, vi + 55 pp, $1.75 (P). [ISBN: 0-486-23895-4] 63 puzzles, new and often surprising, requiring cleverness, ingenuity and only a little bit of school mathematics. Solutions, of course, in the second half. The author is former editor of the British periodical Games and Puzzles. LAS

General. Flatland. Edwin A. Abbott. Arion Pr, 1980, $400. A stunning, creative, limited edition, hand-set and hand-printed republication of Abbott's classic. The volume is printed on 7" x 14" accordion-folded paper panels that open out in a 33 foot-long continuous sheet. The book folds into an aluminum case that opens out for display. Fine printer Andrew Hoyem used die cutting and hand-coloring to illustrate the geometric citizens of Flatland. LAS

Precalculus, T(13: 1). Concepts of Trigonometry. Howard E. Campbell. Prindle, 1981, vii + 305 pp. [ISBN: 0-87150-299-2] Basic skills orientation. Sections reviewing concepts and techniques close each chapter. Some application problems. Angle approach used exclusively. Calculator use explained and encouraged. MW

Education, S(15-16). Math Squared: Graph Paper Activities for Fun and Fundamentals. David Stern. Teachers Coll Pr, 1980, vi + 115 pp, $5.50 (P). [ISBN: 0-8077-2585-4] Enrichment activities for junior high students, ranging from prime numbers through magic squares and figurate numbers to finding slopes. The connecting link is square-ruled graph paper, and readers can play games or solve puzzles as they progress through the material. Useful supplementary reading for preservice teachers looking for new ideas. MW

History, S*(13-18), P, L. Project Whirlwind: The History of a Pioneer Computer. Kent C. Redmond, Thomas M. Smith. Digital Pr, 1980, xiv + 280 pp, $21. [ISBN: 0-932376-09-6] History of the design and construction of the first high speed digital computer able to operate in real time with synchronous parallel logic and the first to use a random access magnetic core memory. Provides the technological and administrative background from which the Project developed. Includes the engineering feats and description of the research and development process over the period 1945-1956 that surrounded the Project. Many photos of historical interest. Appendices. Notes. Index. RJA

History, S, L. From Dits to Bits: A Personal History of the Electronic Computer. Herman Lukoff. Robotics Pr, 1979, xvi + 219 pp, $12.95. [ISBN: 89661-002-0] Reminiscences of one of the pioneers in computer development who, along with John Mauchly and Presper Eckert, was instrumental in developing the early UNIVAC computers. LAS

History, S. Development of Modern Mathematics. J.M. Dubbey. Crane, Russak, 1975, 145 pp, $4.75 (P). [ISBN: 0-8448-0656-0] Reprint of 1975 paperback edition of 1972 edition (TR, August-September 1973). LAS

History, P. The Medieval Latin Translation of the Data of Euclid. Shuntaro Ito. Birkhäuser, 1980, ix + 256 pp, $34. [ISBN: 3-7643-3005-8] Euclid's Data is a study of relations among "givens." For example, Prop. 39: If each side of a triangle is given in magnitude, the triangle is given in species. Ito introduces the first modern English translation (presented in parallel with an edited Latin text) by a comparative analysis of the only four extant Latin manuscripts. LAS

Foundations, P. General Recursion Theory: An Axiomatic Approach. Jens E. Fenstad. Springer-Verlag, 1980, xi + 225 pp, $42.90. [ISBN: 0-387-09349-4] Introductory monograph organized around general theory of computations. Assumes knowledge of Rogers' or comparable text. Four parts cover general theory, finite theories on one and two types, infinite theories, and set recursion and computations in higher types. Epilogue on open problems and areas for further research. KS

Foundations, T*(17: 1), P. Set Theory: An Introduction to Independence Proofs. Kenneth Kunen. Stud. in Logic and Found. of Math., V. 102. North-Holland, 1980, xvi + 313 pp, $29.25. [ISBN: 0-444-85401-0] Introduction to relative consistency proofs in set theory, intended to make the subject accessible to those mathematicians whose research is sensitive to axiomatics. The approach (highlighting Martin's axiom) nicely separates the mathematical and combinatorial techniques from the metamathematical difficulties. No specific training in logic is presupposed, but "maturity" will be called for. Covers much material and has many exercises. GHM

Combinatorics, S(17-18), P. Selected Papers of W.T. Tutte. Ed: D. McCarthy, R.G. Stanton. Charles Babbage Res Centre, 1979. V. I, 438 pp; V. 2, 440 pp, $39.50 (P) set. [ISBN: 0-9690778-0-7] Fifty-

three of the most influential papers from the work (1940-1977) of one of the world's most prolific and profound graph theorists, honoring the author on his 60th birthday. Most interesting are modest introductions to each by Tutte himself, with historical and other insightful perspectives. SS

Number Theory, S(18), P*, L. Old and New Problems and Results in Combinatorial Number Theory. P. Erdős, R.L. Graham. L'Enseignement Mathématique, 1980, 128 pp, Fr. 38 (P). In this monograph the authors discuss various problems in elementary number theory, most of which have a combinatorial flavor. Some are only exercises while others are very difficult or even hopeless. The list of references in this book is a valuable resource. CEC

Algebra, P. Lecture Notes in Mathematics-818: Fixed Rings of Finite Automorphism Groups of Associative Rings. Susan Montgomery. Springer-Verlag, 1980, vii + 126 pp, $9.80 (P). [ISBN: 0-387-10232-9] Expanded version of lectures on results obtained by author and others since 1970. Covers work on chain conditions, polynomial identities, and modules when ring has no $|G|$-torsion or no nilpotent elements or automorphism group contains no generalized inner automorphisms. KS

Algebra, P. Symétries jauges et variétés de groupe. Yuval Ne'eman. Pr U Montreal, 1979, 141 pp, $10 (P). [ISBN: 2-7606-0441-1]

Finite Mathematics, T(13), S. Mathematics, Contemporary Topics and Applications. Howard A. Silver. P-H, 1979, 630 pp, $14.95. [ISBN: 0-13-563304-4] A finite mathematics text for the non-science college student. Topics include problem solving, geometry and graphing, consumer mathematics, probability and statistics. JG

Finite Mathematics, T(13), S. Mathematics: An Everyday Language. Ruric E. Wheeler, Ed R. Wheeler. Wiley, 1979, xi + 483 pp, $15.95. [ISBN: 0-471-03423-1] A text for an introductory college mathematics course emphasizing problem solving using a small collection of mathematical concepts and tools. Includes chapters on flow charts, computers, linear programming, probability and statistics. JG

Finite Mathematics, T(13), S. Finite Mathematics, An Elementary Approach. Lawrence G. Gilligan, Robert B. Nenno. Goodyear, 1977, xiii + 463 pp. [ISBN: 0-87620-327-6] A finite mathematics text with emphasis on probability, matrices, linear programming, game theory, statistics and modeling. JG

Finite Mathematics, T(13), S. Mathematics Is.... Jerome E. Kaufmann. Prindle, 1979, viii + 503 pp, $15.95. [ISBN: 0-87150-263-1] A text for a course in mathematics for the liberal arts student. Topics include logic, consumer mathematics, probability and statistics. There are chapters on calculator and computer mathematics including an introduction to Basic. JG

Calculus, S?. The First Systems of Weighted Differential and Integral Calculus. Jane Grossman, Michael Grossman, Robert Katz. Archimedes Foundation (Box 240, Rockport, MA 01966), 1980, vi + 55 pp, $3 (P). Unconventional development in a specialized notation of an eccentric theory of measure and Stieltjes integration. Like its predecessor Non-Newtonian Calculus (TR, May 1973), the exposition takes place in an intellectual vacuum, with no applications, no references and no relation to other parts of mathematics. LAS

Differential Equations, T(14-15: 1, 2), S, L. Gewöhnliche Differentialgleichungen. V.I. Arnol'd. Springer-Verlag, 1980, 275 pp, $29.20. [ISBN: 0-387-09216-1] An interestingly different first course in ordinary differential equations. Studies the geometry of vector fields and the theory of autonomous systems. Primarily mechanics examples. Very geometric, many pictures, few techniques, no exercises. PH

Differential Equations, S(16-17). Equations of Mathematical Physics. A.V. Bitsadze. Trans: V.M. Volosov, I.G. Volosova. MIR Pub, 1980, 318 pp, $8. Covers the theory of partial linear differential equations and some elements of the theory of linear integral equations. No exercises are included. AO

Differential Equations, S(17-18), P. Applied Methods in the Theory of Nonlinear Oscillations. V.M. Starzhinskii. Trans: V.I. Kisin. MIR Pub, 1980, 263 pp, $8. Exposition of techniques for the analysis of essentially nonlinear autonomous systems of ordinary differential equations. AO

Differential Equations, T*(17-18: 1), S, P, L.** Quadratic Form Theory and Differential Equations. John Gregory. Math. in Sci. and Eng., V. 152. Acad Pr, 1980, xii + 237 pp, $29.50. [ISBN: 0-12-301450-6] Unified theory of quadratic forms including a new formal theory of approximation of quadratic forms/linear operators on Hilbert spaces. Qualitative comparison results such as generalized Sturm separation theorems of differential equations. Approximation theory applicable to easily implemented numerical algorithms. Much seed material and many unanswered problems for further development. Connections with variational problems, eigenvalue problems, Sturm-Liouville boundary-value problems, spline approximations and singular differential equations in mathematical physics. JK

Differential Equations, T*(14: 1, 2). Applied Differential Equations, Third Edition. Murray R. Spiegel. P-H, 1981, xvi + 717 pp, $21.95. [ISBN: 0-13-040097-1] Ordinary differential equations, systems, and partial differential equations. Expanded third edition includes new topics (non-linear differential equations and stability, matrix methods, some special functions), a larger number and a wider variety of applications, and a more flexible format. Suitable for one- or two-semester

courses. Writing style is clear and concise. Plenty of exercises--some straightforward, some challenging. Excellent choice for sophomore level courses. (Second Edition, TR, October 1967.) JK

Differential Equations, T(17-18: 1, 2), P. Nonlinear Oscillations. Ali Hasan Nayfeh, Dean T. Mook. Wiley, 1979, xiv + 704 pp, $38.50. [ISBN: 0-471-03555-6] Presents the state-of-the-art in the theory of single-frequency excitations and single-degree-of-freedom systems, as well as multiharmonic excitations and systems with several degrees of freedom. Very extensive bibliography. AO

Numerical Analysis, S(15-17), P. Software for Roundoff Analysis of Matrix Algorithms. Webb Miller, Celia Wrathall. Comp. Sci. and Appl. Math. Acad Pr, 1980, x + 151 pp, $18.50. [ISBN: 0-12-497250-0] Presents techniques for analyzing the propagation of rounding errors in matrix algorithms, and also serves as the documentation for a software package implementing these techniques (ACM Algorithm 532). AO

Numerical Analysis, S(15), P, L. The Calculus of Finite Differences. L.M. Milne-Thomson. Chelsea, 1981, xxiv + 558 pp, $22.50. A second, textually unaltered edition of a work originally published in 1933. This edition is printed on special 'long-life' acid-free paper. CEC

Functional Analysis, P. Direct Integral Theory. Ole A. Nielsen. Lect. Notes in Pure and Appl. Math., V. 61. Dekker, 1980, viii + 165 pp, $23.50 (P). [ISBN: 0-8247-6971-6] Thorough study of the direct integral theory of von Neumann algebras acting on separable Hilbert spaces and of representations on separable Hilbert spaces of either separable involutive Banach algebras or separable locally compact groups. Employs Borel structure and Effros' approach to direct integrals. TRS

Functional Analysis, P. Integral Equations with Fixed Singularities. Roland Duduchava. Teubner, 1979, 172 pp, (P). Devoted to the study of equations with fixed singularities in the kernels, this monograph is based on the theory of non-classical integral equations in convolution on the semi-axis. The Noetherian properties, the index and the solutions of these equations are investigated in Lebesgue spaces; in particular, asymptotic behavior near end points and smoothness inside the interval of solution are studied. TRS

Functional Analysis, P. Operator Theory and Functional Analysis. Ed: I. Erdelyi. Pitman, 1979, 164 pp, $15 (P). [ISBN: 0-8224-8450-1] The ten invited addresses of a special session of the 1978 AMS summer meeting in Providence. Papers on the invariant subspace problem, spectral theory, Wold decompositions, generalized Calkin algebras, peak functions, and applications of functional analysis to topological measure theory. TRS

Optimization, P. Reduction Methods in Nonlinear Programming. G. van der Hoek. Math. Centre Tracts, No. 126. Math Centrum, 1980, v + 194 pp, Dfl. 24 (P). [ISBN: 90-6196-199-8] Concerned with programming problems in which some of the decision variables appear in a nonlinear way in the objective function or in one or more of the constraints. The approach calls for reducing the degree of nonlinearity and/or the number of constraints. Photocopied from typed copy of research done at the Mathematical Center in Amsterdam. AWR

Optimization, T(15-16: 1), L. Optimisation. D.M. Greig. Longman, 1980, xii + 179 pp, $14.50 (P). [ISBN: 0-582-44186-2] Concise account of theoretical results in constrained and unconstrained optimization. Applications of linear programming to transportation problems, game theory, and allocation of resources. Some exercises supplement the text. JRG

Optimization, T, P, L. Location on Networks: Theory and Algorithms. Gabriel Y. Handler, Pitu B. Mirchandani. MIT Pr, 1979, xviii + 233 pp, $25. [ISBN: 0-262-08090-7] A text on algorithmic solutions to network location problems which merges the topics of network analysis and location theory. No exercises. JG

Analysis, P. Introduction to Pseudodifferential and Fourier Integral Operators. François Treves. Plenum Pr, 1980. Volume 1, Pseudodifferential Operators, xxxix + 299 pp, $29.95 [ISBN: 0-306-40403-6]; Volume 2, Fourier Integral Operators, xxv + 348 pp, $35. [ISBN: 0-306-40404-4] Two high-power volumes that demand, because of both content and the style of the author, a strong background in real and complex analysis. AWR

Analysis, T(17), S, P, L. Continued Fractions: Analytic Theory and Applications. William B. Jones, W.J. Thron. Ency. Math. and its Appl., V. 11. A-W, 1980, xxviii + 428 pp, $37.50. [ISBN: 0-201-13510-8] The first comprehensive and self-contained exposition of the analytic theory of continued fractions in over twenty years. The emphasis is on applications and computational methods. Assumes a knowledge of complex analysis. No exercises. Includes an excellent bibliography. CEC

Analysis, T(16-17), L. Introduction to Perturbation Techniques. Ali Hasan Nayfeh. Wiley, 1981, xiv + 519 pp, $29.95. [ISBN: 0-471-08033-0] An introduction to asymptotic expansions using examples primarily from algebraic and ordinary differential equations. Material includes: asymptotic expansions of integrals, determination of the adjoints of homogeneous linear equations, solvability conditions of linear inhomogeneous problems, the Mathieu equation, a summary of trigonometric identities, and properties of linear ordinary differential equations. Chapter exercises. TRS

Algebraic Geometry, T(17), S. Plane Algebraic Curves: An Introduction Via Valuations. Grace Orzech, Morris Orzech. Pure and Appl. Math., V. 61. Dekker, 1981, viii + 225 pp, $29.50. [ISBN: 0-8247-1159-9] A down-to-earth introduction suitable for advanced undergraduates and beginning graduates.

Important topics are: smoothness, division, genus, and Riemann-Roch. Less ambitious than Fulton's text. Many exercises. SG

Topology, T(18: 1, 2), S, P. Some Applications of Topological K-Theory. N. Mahammed, R. Piccinini, U. Suter. Math. Stud., No. 45. North-Holland, 1980, ix + 317 pp, $41.50 (P). [ISBN: 0-444-86113-0] A K-theory retrospective which presents a readable and systematic account of the development and most famous applications of K-theory (e.g., vector fields on spheres and the Atiyah-Singer Index theorem). The text provides both flavor and fact to enlighten the reader. JAS

Topology, T*(16-18: 1), L. Singular Homology Theory. William S. Massey. Grad. Texts in Math., No. 70. Springer-Verlag, 1980, xii + 265 pp, $24.80. [ISBN: 0-387-90456-5] A cubical treatment of singular homology appropriate for a second course in algebraic topology. The approach reflects experience with presenting the topic over the past couple of decades and includes a nice historical and motivational first chapter. JAS

Probability, T(17), S, P. Finite Markov Processes and Their Applications. Marius Iosifescu. Wiley, 1980, 295 pp, $32.50. [ISBN: 0-471-27677-4] A revised and expanded version of the 1977 edition (in Romanian). An essentially complete treatment of finite Markov chains at a level accessible to a novice in the field. An unusual feature is the treatment of inhomogeneous chains, which are currently of growing interest and importance. An extensive bibliography, emphasizing Romanian contributors to Markov processes. TAV

Probability, P. Applied Stochastic Processes. Ed: G. Adomian. Acad Pr, 1980, ix + 301 pp, $21. [ISBN: 0-12-044380-5] Proceedings of a conference held May 1978 at the Center for Applied Mathematics, Athens, Georgia. The common thread of the twelve papers contained in this book seems to be real solutions to stochastic models of real phenomena. TAV

Probability, S(15), P. Essay D'Analyse sur les Jeux de Hazard, Third Edition. Pierre Remond de Montmort. Chelsea, 1980, xlii + 414 pp. [ISBN: 0-8284-0307-4] Photographic reprint of the second edition of 1713. Analyzes various games of chance popular in 18th century Europe, e.g., piquet, bassete, and others. Concludes with correspondence between the author and Nicolas Bernoulli. JG

Probability, S(13-18), P, L. Studies in Subjective Probability. Ed: Henry E. Kyburg, Jr., Howard E. Smokler. Krieger, 1980, 262 pp, $10.50 (P). [ISBN: 0-88275-296-0] Selected papers by Ramsey, de Finetti, Koopman, I.J. Good, L.J. Savage, and Jeffrey. Extensive bibliography. Reprint of 1964 Wiley publication. FLW

Statistics, T(16-18). Parameter Estimation: Principles and Problems. Harold W. Sorenson. Control and Systems Theory, V. 9. Dekker, 1980, xi + 382 pp, $45. [ISBN: 0-8247-6987-2] Classical estimation theory with reviews in appendices of matrix algebra and probability theory. Nothing on ridge regression or Stein estimation. FLW

Statistics, T(16-18: 1), P. Approximation Theorems of Mathematical Statistics. Robert J. Serfling. Wiley, 1980, xiv + 371 pp, $34.95. [ISBN: 0-471-02403-1] Covers a broad range of limit theorems useful in mathematical statistics. FLW

Statistics, T*(15-16: 1, 2). An Introduction to Mathematical Statistics and its Applications. Richard J. Larsen, Morris L. Marx. P-H, 1981, xii + 596 pp, $22.95. [ISBN: 0-13-487744-6] Well-written introduction to standard topics, at a level presupposing three semesters of calculus. Suitable for either terminal or preparatory courses, it emphasizes inter-relationships between probability, statistics and data analysis. Extensive historical notes and case studies. Modest departure from conventional ordering of topics for well considered pedagogical reasons. "Questions" scattered strategically throughout text, with many review exercises at chapter ends. Appears to be a very "teachable" text, with no loss of rigor. GHM

Statistics, T(15-17: 1), L. Regression: A Second Course in Statistics. Thomas H. Wonnacott, Ronald J. Wonnacott. Wiley, 1981, xix + 556 pp, $26.95. [ISBN: 0-471-95974-X] In the Wiley Series in Probability and Mathematical Statistics. Chapters 2-10 and 12-15 are essentially identical to the same chapters in their 1979 text Econometrics, Second Edition (TR, March 1980). This text contains a different introductory chapter, a different Chapter 11 on analysis of variance, and new sections on path analysis and spectral analysis. RSK

Statistics, T(18: 1), S, P*. Matrix Derivatives. Gerald S. Rogers. Lect. Notes in Stat., V. 2. Dekker, 1980, v + 209 pp, $27.50 (P). [ISBN: 0-8247-1176-9] Comprehensive survey of matrix derivatives in statistics, requiring a background in advanced calculus, mathematical statistics and matrix algebra. Material is well-documented and there is a good bibliography. No exercises. RSK

Statistics, P. Multivariate Analysis, Second Edition. Maurice Kendall. Macmillan, 1980, 210 pp, $29.95. [ISBN: 0-02-847790-1] Minor revision of the author's 1975 First Edition. Topics covered include principal components, classification and clustering, factor analysis, canonical correlations, some distribution theory, problems in regression analysis, functional relationship, tests of hypotheses, discrimination, and categorized multivariate data. RSK

Statistics, T(13-14: 1), S, L. Ideas of Statistics. J. Leroy Folks. Wiley, 1981, xiii + 368 pp, $17.95. [ISBN: 0-471-02099-0] An elementary treatment of the basic ideas with many historical notes and additional short chapters on some of the controversies. FLW

Statistics, P*. Analysis of Variance. Ed: P.R. Krishnaiah. Handbook of Stat., V. 1. North-Holland, 1980, xvii + 1002 pp, $134.25. [ISBN: 0-444-85335-9] First volume of a new series of reference books in statistical methodology and applications, each volume to be devoted to a particular topic. Contains twenty-five chapters written by prominent statisticians and covers most of the useful techniques in univariate and multivariate ANOVA. Material is primarily expository, but technical, and well-documented. RSK

Statistics, T(16-17: 1), S, P, L. Order Statistics, Second Edition. H.A. David. Wiley, 1981, xiii + 360 pp, $34.95. [ISBN: 0-471-02723-5] In the Wiley Series in Probability and Mathematical Statistics. Revision of the author's 1970 First Edition (TR, February 1971). Contains new sections on order statistics for independent nonidentically distributed variates, on linear functions of order statistics, on concommitants of order statistics, and on testing for outliers from a regression model, plus an expanded section on robust estimation and rewritten sections on asymptotic theory. Bibliography has also been expanded and updated. RSK

Statistics, T(16-17: 2). Probability Theory and Mathematical Statistics, Third Edition. Marek Fisz. Krieger, 1980, xvi + 677 pp, $34.50. [ISBN: 0-89874-179-3] Reprint with corrections of the author's 1963 Third Edition. Contents are roughly half probability theory, including Markov chains and stochastic processes, and half mathematical statistics, including sequential analysis, but little regression analysis. RSK

Computer Programming, T(15-17: 1, 2), P, L. Functional Programming: Application and Implementation. Peter Henderson. P-H, 1980, xi + 355 pp, $33.95. [ISBN: 0-13-331579-7] Claims to give complete coverage of the functional, or applicative, style of programming which is increasingly important in the study of semantics of programming languages and in artificial intelligence applications. Discusses structure and general principles at a reasonably theoretical level, but also stresses applications and implementation. Includes a compiler and instructions for implementing a purely functional language Lispkit Lisp so that concepts discussed in the text can be experienced concretely. Includes exercises. GHM

Computer Programming, T(13-16: 1). Problem Solving and Structured Programming in Pascal. Elliot B. Koffman. A-W, 1981, xiv + 483 pp, $$13.95 (P) [ISBN: 0-201-03893-5] Emphasis on good programming habits through the technique of stepwise algorithm development. Moves gradually through large amounts of material on Pascal, getting up to more advanced data structures in last three chapters. Problem solving skills are instilled through a series of carefully structured solved examples, emphasizing three basic phases: analysis of the problem, stepwise specification of algorithm, and implementation in Pascal. Moderate number of exercises. Presupposes no programming experience. GHM

Computer Programming, T(8-14), S. Are You Computer Literate? Karen Billings, David Moursund. Dilithium Pr, 1979, viii + 148 pp, $6.95 (P). [ISBN: 0-918398-29-0] General discussion of the nature and role of computers. Each chapter begins with simple-minded True-False self-quiz. Includes a good reference list, a glossary, and a concluding multiple-choice computer literacy exam. LAS

Computer Science, S(16-18), P. Lecture Notes in Computer Science-91: Grammar and L Forms: An Introduction. Derick Wood. Springer-Verlag, 1980, ix + 314 pp, $19.50 (P). [ISBN: 0-387-10233-7] Goal is to provide an introductory, unified account of context-free form theory. Begins with basic terminology, notation, conventions and with a survey of approaches to grammatical similarity. Includes context-free grammar forms, EOL and ETOL forms, some results for non-context-free forms, and concludes with a short chapter on historical background and some open problems. Extensive bibliography. RJA

Computer Science, P. Lecture Notes in Computer Science-93: Context-Free Grammars: Covers, Normal Forms, and Parsing. Anton Nijholt. Springer-Verlag, 1980, vii + 253 pp, $16.80 (P). [ISBN: 0-387-10245-0] Main emphasis of this monograph is on algorithmic transformations of context-free grammars to context-free grammars in some normal form or to other grammars which have useful parsing properties. A quite general "grammar cover" approach is used, in competition with the "grammar functor" approach. Sizable bibliography. GHM

Computer Science, P. Lecture Notes in Computer Science-90: Using Sophisticated Models in Resolution Theorem Proving. David M. Sandford. Springer-Verlag, 1980, xi + 239 pp, $14 (P). [ISBN: 0-387-10231-0] For the reader already acquainted with the resolution method of automatic theorem proving, this monograph describes a major strategy for improving the efficiency of such methods. The Hereditary Lock Resolution method employs both a syntactic and a semantic component to reduce redundancy and speed the search for a proof. GHM

Computer Science, S(15-17), P, L*. Structured System Programming. Jim Welsh, Michael McKeag. P-H, 1980, xii + 324 pp, $24.95. [ISBN: 0-13-854562-6] Presents two case studies illustrating the application of structured programming to large systems programs. Includes as an intergral part of the text complete listings of a compiler and an operating system, both written in Pascal Plus. AO

Systems Theory, S(15-18), P, L. Fuzzy Sets and Systems: Theory and Applications. Didier Dubois, Henri Prade. Math. in Sci. and Eng., V. 144. Acad Pr, 1980, xvii + 393 pp, $49.50. [ISBN: 0-12-222750-6] In the foreword L.A. Zadeh describes this as "a comprehensible research monograph that covers almost all of the important developments in the theory of fuzzy sets and in their applications that have taken place during the past several years." FLW

Systems Theory, P, L. Search Theory and Applications. Ed: K. Brian Haley, Lawrence D. Stone. Plenum Pr, 1980, ix + 277 pp, $35. [ISBN: 0-306-40562-8] Collection of papers given at the 1979 NATO Advanced Research Institute. Survey of search theory and discussion of applications to such areas as search and rescue operations, exploration, and surveillance. JRG

Systems Theory. Algebraic and Geometric Methods in Linear Systems Theory. Ed: Christopher I. Byrnes, Clyde F. Martin. Lect. in Appl. Math., V. 18. AMS, 1980, viii + 327 pp, $12.40. [ISBN: 0-8218-1118-5] Papers from the June 1979 seminar held at Harvard University. JAS

Applications (Computer Hardware), T(15-18: 1, 2), S, P, L. Computer Engineering: A DEC View of Hardware Systems Design. C. Gordon Bell, J. Craig Mudge, John E.McNamara. Digital Pr, 1978, xxii + 585 pp, $19.95. [ISBN: 0-932376-00-2] Employs the case study approach. Uses four families of DEC computers (PDP-1, PDP-8, PDP-11, PDP-10) to illustrate in detail the design of machines that implement instruction sets. Emphasizes lower level technological, economic, organizational, and environmental forces affecting the evolution of these computer families. Special emphasis is placed on computer modules and their evolution. Appendices. Bibliography. Index. RJA

Applications (Cosmology), P. The Large-Scale Structure of the Universe. P.J.E. Peebles. Princeton U Pr, 1980, xiii + 422 pp, $30; $9.95 (P). [ISBN: 0-691-08239-1; 0-691-08240-5] A monograph that focuses on the large-scale distribution of matter throughout the universe. Empirical data on the clustering of galaxies is analyzed as an aid to understanding the nature and evolution of the universe. AO

Applications (Electrical Engineering), T(17). Active Network and Feedback Amplifier Theory. Wai-Kai Chen. Hemisphere Pub, 1980, xiii + 481 pp, $27.50. [ISBN: 0-07-010779-3] A textbook in network theory for first-year graduate students in electrical engineering. Emphasizes matrix techniques. AO

Applications (Engineering), S(17-18), P. Physical Applications of Stationary Time-Series with Special Reference to Digital Data Processing of Seismic Signals. Enders A. Robinson. Macmillan, 1980, xi + 302 pp, $42. [ISBN: 0-02-851050-X] A detailed exposition of the theory of stationary time series used in seismic engineering. AO

Applications (Physics), S(15-17), P. Arithmetic Applied Mathematics. Donald Greenspan. Pergamon Pr, 1980, vii + 165 pp, $11.25 (P); $25. [ISBN: 0-08-025046-7; 0-08-025047-5] A reformulation of Newtonian and special relativistic mechanics using only arithmetic is presented at a level appropriate for undergraduates. The theoretical background is presented together with the results of model calculations for several types of complex physical phenomena. AO

Applications (Physics), T(14-15: 1). Continuum Mechanics. A.J.M. Spencer. Longman, 1980, 183 pp, $13.50 (P). [ISBN: 0-582-44282-6] The basic principles of continuum mechanics are presented at a level appropriate for undergraduate study. Introductory chapters cover material on matrix algebra, vectors and tensors used in the remainder of the book. AO

Applications (Physics), T(15-16: 1). An Introduction to the Theory of Elasticity. R.J. Atkin, N. Fox. Longman, 1980, 245 pp, $14.95 (P). [ISBN: 0-582-44283-4] An elementary introduction (at the undergraduate level) to the finite and infinitesimal theory of elasticity. Introductory material on the general principles of continuum mechanics makes the text relatively self-contained. AO

Applications (Physics), S(16-18), P, L. Essays in General Relativity: A Festschrift for Abraham Taub. Ed: Frank J. Tipler. Acad Pr, 1980, xviii + 236 pp, $30. [ISBN: 0-12-691380-3] A collection of essays to honor Abraham Taub on the occasion of his retirement from the mathematics faculty of the University of California at Berkeley. Some of these papers are from the August 1978 Berkeley Symposium held in honor of Taub. Although the papers are in a sense expository they represent as well "state-of-the-art" scholarly and philosophical ideas, e.g., Tipler's survey of ancient Babylonians, et al., eternal return idea and his proof of the no-return theorem. JAS

Applications (Social Science), P. Mathematical Frontiers of the Social and Policy Sciences. Ed: Loren Cobb, Robert M. Thrall. Westview Pr, 1981, xiv + 186 pp, $22. [ISBN: 0-89158-953-8] Five papers from the 1979 AAAS annual meeting on cognitive psychology, stochastic differential equations, soft data analysis, multicriterion decision analysis and a multiattribute utility analysis of the Los Angeles school desegregation options. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036

PERSONAL ITEMS

Sterling Berberian of the University of Texas at Austin will be visiting the University of Reading during the fall term of 1981, and the University of Poitiers during the first half of 1982.

Harold Don Allen of Nova Scotia Teachers College has been named Editor of *The Mathematical Log,* Mu Alpha Theta publication, for a three-year term.

Mildred Reigh, a member of the mathematics faculty at Indiana University of Pennsylvania since 1963, will be retiring in June, 1981.

Philip Caverly, Associate Professor of Mathematics, has been appointed Associate Professor of Computer Science and Chairman of the Computer Science Department at Jersey City State College.

Albert Wilansky of Lehigh University served as Visiting Professor of Mathematics at Universität Bern for the month of April, 1981, and at Tel Aviv University for the month of May.

At Youngstown State University, *John J. Buoni* and *J. Douglas Faires* have been promoted to Professor.

C. Ray Crain is now Manager of Statistics at McNeil Pharmaceutical of Spring House, PA.

William H. Bradford of Winnfield, LA, died Feb. 2, 1981, at the age of 68. He was a member of the Association for 39 years and had served a term as Chairman of the Louisiana-Mississippi Section.

Isidore Dressler, most recently an adjunct member of the faculty of Pace University, died Dec. 23, 1980, at the age of 72. He was a member of the Association for 12 years.

Harry M. Gehman, for many years Secretary-Treasurer of MAA, died in January 1981 at Los Gatos, California. A full account of his life and his work for the Association will appear in a future issue of *The Monthly.*

Jack U. Russell, Professor of Mathematics at Southwestern at Memphis, died Feb. 14, 1981, at the age of 53. He was a member of the Association for 27 years.

The death of *Henry Scheffe* of the Department of Mathematics of Indiana University has been reported. He was a member of the Association for 40 years.

COMING SOON IN THE MONTHLY

The following articles are among those which the editors have accepted for future issues of the MONTHLY. The order of listing does not indicate the order in which they will appear.

William Abikoff, *The Uniformization Theorem*
Henry L. Alder, *Harry Merrill Gehman 1898-1981*
W. Brian Arthur, *Why a Population Converges to Stability*
R.P. Boas, *Can We Make Mathematics Intelligible?*
J.W. Bruce, P.J. Giblin, and C.G. Gibson, *On Caustics of Plane Curves*
Roger L. Cooke, *Almost-Periodic Functions*
Paul W. Dixon, *Transfinite Solution to Last Theorem of Fermat*
Josef Dodziuk, *Eigenvalues of the Laplacian and the Heat Equation*
L.E. Dubins and D.A. Freedman, *Machiavelli and the Gale-Shapley Algorithm*
A.J.W. Duijvestijn, P.J. Federico and P. Leeuw, *Compound Perfect Squares*
Solomon W. Golomb, *Normed Division Domains*
Solomon W. Golomb, *Irrational Sums and Twin Primes*
I. Grattan-Guiness, *On the Development of Logics Between the Two World Wars*
Eldon Hansen, *Sums of Functions Satisfying Recursion Relations*
M. Katchalski, M.S. Klamkin and A. Liu, *An Experience in Problem Solving*
Patricia Kenschaft, *Black Women in Mathematics in the United States*
Saunders MacLane, *Mathematical Models--a Sketch for the Philosophy of Mathematics*
Katherine E. McLain and Hugh M. Edgar, *A Note on Golomb's "Cyclotomic Polynomials and Factorization Theorems"*
Lawrence Narici and Edward Beckenstein, *Strange Terrain--Nonarchimedean Spaces*
Anthony Ralston, *Computer Science, Mathematics and the Undergraduate Curricula in Both*
Karen D. Rappaport, *S. Kovalevsky*
Bhana Srinivasan, *Characters of Finite Groups: Some Uses and Mathematical Applications*
Wolfgang Walter, *A New Approach to Euler's Trigonometric Expansions*
Edward T. Wong, *Polygons, Circulant Matrices, and Moore-Penrose Inverses*

FUTURE PLANS FOR "NEWS AND NOTICES"

The "News and Notices" section of the *Monthly* will be transferred to *MAA Focus,* the newsletter of the Association, according to this schedule:

Announcements - September 1981

Personal Items - January 1982

"Official Reports and Communications" will contine to be published in the *Monthly* as a permanent record of the actions, recommendations, and reports of the Association. It is hoped that these changes will improve communications within the Association and avoid unnecessary duplication.

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

NEW SECTIONAL GOVERNORS OF THE ASSOCIATION

The Mathematical Association of America is pleased to announce the election of the following Governors of the Association representing the Sections indicated:

Allegheny Mountain	Charles A. Cable, Allegheny College
Indiana	Meyer Jerison, Purdue University
Kentucky	Martha F. Watson, Western Kentucky University
Metropolitan New York	Donald J. McCarthy, St. John's University
Nebraska	Gerald W. Johnson, University of Nebraska at Lincoln
Northern California	Jean J. Pedersen, University of Santa Clara
Oklahoma-Arkansas	James H. Yates, Central State University
Rocky Mountain	John H. Hodges, University of Colorado
Wisconsin	Norbert J. Kuenzi, University of Wisconsin at Oshkosh

In accordance with the bylaws of The Association, these Governors were elected by vote of the members of their Sections, in elections conducted by the MAA Headquarters. The two Sections with the highest voter participation were the Nebraska Section with 38.5% voting and the Kentucky Section with 33.1% voting.

Alfred B. Willcox, Executive Director

FEBRUARY 1981 MEETING OF THE LOUISIANA-MISSISSIPPI SECTION OF MAA

The 58th annual meeting of the Louisiana-Mississippi Section of the MAA was held at Mississippi State University, Mississippi State, MS on February 13 and 14, 1981. Mississippi State University served as host for the meeting. There were 168 registered participants for this meeting which was held jointly with the Louisiana-Mississippi Branch of NCTM and the Mississippi Council of Teachers of Mathematics.

Chairman *Edwin P. Oxford* of the University of Southern Mississippi presided. Dr. *Richard D. Anderson,* President of the MAA, gave two invited addresses: "Elementary Topological Properties of Infinite Dimensional Spaces" and "New Thrusts in Mathematical Education." Four panel discussions were held as follows: "High School-College Interface," moderator Prof. *Robert A. Shive;* "The Graduate Training Needs of Two-Year-College Teachers of Mathematics," moderator Prof. *Shelby Harris;* "College Remediation," moderator Prof. *Steve Doblin,* University of Southern Mississippi; "Strengthening Our Master's Degree Program," moderator Prof. *A.J. Hulin,* University of New Orleans. Two microcomputer workshops were also held: "Remedial Mathematics and Microcomputer" with Profs. *T. Flaherty* and *A. Lopez* of Loyola University and "Microcomputers and Calculus Plus" with Profs. *J. Abbott,* University of New Orleans, *David Cook,* University of Mississippi, and *P. Ohme,* Northeast Lousiana University.

Student papers were given as follows: "The Hyperbolic Paraboloid as a Solar Collector," *J.R. Chase;* "Counting the Petals of Rose Curves," *E.A. O'Neal* (these two from the University of Southern Mississippi); "Mathematical Proof and Applications of Tchebysheff's Theorem," *S. Schof;* "Calculating the Normal Probability Density Functions on Microcomputers," *B.S. Henling* (these two from Loyola University); and "Math Anxiety in American Schools," *R. Newell,* Mississippi State University.

The following thirteen papers were contributed: "A Finite Element Model for Simulation of Head and Neck Injury," *Ronald Hosey,* Loyola University; "A Microcomputer *LOGO* Interpreter," *Terry Flaherty,* Loyola University; "Educational Help: High School--University Cooperation in Computer Literacy," *Antonio Lopez,* Loyola University: "A Characterization of 2-Central Groups," *Gary Walls,* University of Southern Mississippi; "On Integral Generalized Inverses," *Jimmy Gilbert,* Louisiana Tech University; "Operator Algebras Generated by Closed Operation on a Separable Hilbert Space," *Clark Rhodes,* Loyola University; "Induced Forests in Cubic Graphs," *William Staton,* University of Mississippi; "External Bipartite Subgroups of Cubic Triangle-Free Graphs," *Glenn Hopkins,* University of Mississippi; "Asymptotic Behavior of the Solutions of a Perturbed Nonlinear Differential Equation," *Paul Spikes* and *Wanda Smith,* Mississippi State University; "Some Results on the Behavior of Oscillatory Solutions of Second Order Nonlinear Differential Equations," *John R. Graef* and *Paul Spikes,* Mississippi State University; "Primary Extensors and Time Dependent Functions," *Gertrude Okhuysen,* Mississippi State University; and "Contractive Mappings and Elementary Numerical Analysis," *Jimmy L. Solomon,* Mississippi State University.

In the business meeting the following officers were elected for 1981-82: Chairman, *Adam J. Hulin,* University of New Orleans; Louisiana Vice-Chairman, *Duane Blumberg,* University of Southwestern Louisiana; Mississippi Vice-Chairman, *Paul Spikes,* Mississippi State University; Secretary-Treasurer and Newsletter Editor (3 year term), *J.L. Tilley,* Mississippi State University.

SUGGESTION BOX

Members of the MAA are encouraged to send in suggestions, questions, etc., about the operation of the Association. Communications will be referred to the appropriate officer of the Association for answering; from time to time those of general interest may also be answered in one or both of the official journals. Communications should be addressed to: Suggestion Box, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington D.C. 20036.

CALENDAR OF FUTURE MEETINGS

Sixty-first Summer Meeting, University of Pittsburgh, Pittsburgh, Pennsylvania, August 17–19, 1981.
Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, last weekend in April or first weekend in May. Deadline for papers six weeks before meeting.
EASTERN PENNSYLVANIA AND DELAWARE, Saturday before Thanksgiving.
FLORIDA, early March. Deadline for paper titles two weeks before meeting.
ILLINOIS, first Friday/Saturday in May.
INDIANA
INTERMOUNTAIN
IOWA, third weekend in April. Deadline for papers February 1.
KANSAS, March or April. Deadline for papers January 1.
KENTUCKY, early April. Deadline for papers six weeks before meeting.
LOUISIANA-MISSISSIPPI, Friday-Saturday before February 20. Deadline for papers three months before meeting.
MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA, Saturday before Thanksgiving and last Saturday in April.
METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.
MICHIGAN, first Friday and Saturday in May. Deadline for papers six weeks before meeting.
MISSOURI, late March/early April. Deadline for papers January 31.
NEBRASKA, April.
NEW JERSEY, early November and early May.
NORTH CENTRAL, end of October and April. Deadline for papers October 1 and April 1.
NORTHEASTERN, New England College, Henniker, New Hampshire, June 12–13, 1981.
NORTHERN CALIFORNIA, first or second Saturday in February.
OHIO
OKLAHOMA-ARKANSAS, (approx.) Friday and Saturday of first weekend in April. Deadline for papers three weeks before meeting.
PACIFIC NORTHWEST, Lewis and Clark College, Portland, Oregon, June 19–20, 1981.
ROCKY MOUNTAIN, last weekend in April or first weekend in May. Deadline for papers eight weeks before meeting.
SEAWAY, first Saturday in November and Saturday in late April. Deadline for papers six weeks before meeting.
SOUTHEASTERN
SOUTHERN CALIFORNIA, first or second Saturday in March.
SOUTHWESTERN, usually in April. Deadline for papers two weeks before meeting.
TEXAS, Friday and Saturday in early April. Deadline for papers March 1.
WISCONSIN, Friday and Saturday between mid-April and first week in May. Deadline for papers six weeks before meeting.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES
AMERICAN MATHEMATICAL SOCIETY, University of Pittsburgh, Pittsburgh, Pennsylvania, August 18–21, 1981.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION, Los Angeles, California, June 22–25, 1981.
ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.
ASSOCIATION FOR SYMBOLIC LOGIC
ASSOCIATION FOR WOMEN IN MATHEMATICS, University of Pittsburgh, Pittsburgh, Pennsylvania, August 17–21, 1981.
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA, University of California, Los Angeles, August 9–12, 1981.
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, April 14–17, 1982.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Regency Hyatt House, Houston, Texas, October 12–14, 1981.
PI MU EPSILON, University of Pittsburgh, Pittsburgh, Pennsylvania, August 17–20, 1981.
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Troy, New York, June 8–10, 1981.

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

AUGUST - SEPT

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 7

Mathematical Models
by Saunders Mac Lane

Computer Science and Mathematics

Machiavelli Applies to College

Logic Between the Two World Wars

Polygons, Circulants, Moore-Penrose Inverses

Almost-Periodic Functions

$$x_{k+1}=x_k-\frac{f(x_k)}{f'(x_k)-\frac{1}{2}f''(x_k)\dfrac{f(x_k)}{f'(x_k)}}$$ (see p. 530)

Detailed contents on cover 4

1981

THE AMERICAN
MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 7**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are $25 for each of the first two years of membership and $30 thereafter. Student membership is available with annual dues of $15. For nonmembers the subscription price is $36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.
Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

HARRY MERRILL GEHMAN
1898–1981

The Mathematical Association of America suffered a profound loss when on January 15, 1981, Harry Gehman—Mr. MAA, as he was so frequently referred to—passed away on his eighty-third birthday.

He leaves behind as his legacy an organization which he, through his dynamic strength and devoted and diligent service, brought to its present position of effectiveness, efficiency and influence.

When he became the first Chairman (1940–41) of the Upper New York State Section of the Association, he made it his goal both to assist the Sections of the Association as the "grass roots" of the national organization and to strengthen the national organization so that it could effectively carry out its mission at the national level.

It is no surprise then that he was elected to many key positions at the national level, first to the Finance Committee in 1944, later as Secretary-Treasurer in 1948, succeeding Walter B. Carver. It was in this position as Secretary-Treasurer that he was able, through effective, selfless service, to have the most profound influence by vastly expanding MAA's services to its membership and the mathematical community in general.

In the late 1950's, it became necessary to separate the offices of Secretary and Treasurer. Harry remained Treasurer, and I was elected by the Board of Governors to succeed Harry as Secretary as of 1960. Thus, I had the enormous pleasure of becoming well acquainted with Harry, a friendship which I have treasured ever since. The zeal, devotion, thoughtfulness, and thoroughness with which he carried out all his responsibilities were awe-inspiring. He has served as a shining model for many and certainly for me.

Harry continued as Treasurer of the Association, thereby retaining the lion's share of his responsibilities as Secretary-Treasurer. He maintained the office of the MAA with such modest expenditures that its dues remained one of the lowest—if not the lowest—of any comparable organization. When it became increasingly obvious that, in addition to carrying out the duties of Treasurer, he was, in fact, also carrying out those commonly associated with an Executive Director, he was also given the latter title, thus becoming the MAA's first Executive Director, in which position he served until 1968, the year in which he retired from the State University of New York at Buffalo.

Harry's enormous contributions, not only to the MAA but to the mathematical community in general, were organized in 1966, when he was named the recipient of the MAA's Award for Distinguished Service to Mathematics. The citation which Albert Tucker wrote on that occasion appears in the January 1966 issue of this MONTHLY, pages 1–2. This article gives an excellent account of Harry's many contributions up to that time, among which particular note should be made of his active research in the topology of continuous functions (during the period 1925–35, he published 17 papers in the leading American mathematical journals and guided the dissertation work of half a dozen Ph.D. candidates within a period of less than five years) and his service from 1929 to 1962 as Chairman of the Department of Mathematics at the University of Buffalo, now the State University of New York at Buffalo. As Professor Lewis Coburn, the present Chairman of the Department of Mathematics, noted after learning of Harry's death: "Dr. Gehman's distinguished research level can still serve as a bench mark for our junior colleagues. His administrative longevity must be regarded with awe by academic bureaucrats both here and elsewhere." There was an interlude in his service to the University of Buffalo in 1945, when, at a moment's notice, he went to England to head the Department of Mathematics at the Shrivenham University of the United States Army. A colleague of Harry's at Shrivenham reports that Harry personally organized the Department of Mathematics at Shrivenham and that it was probably the best department in the Army university there. Harry personally recruited his staff by telephoning his friends and got many acceptances out of loyalty to him.

Harry moved to Los Gatos, California, in the early 1970's after the untimely death of his wife, Marian, whose devotion to mathematics rivaled that of Harry. He chose Los Gatos since that was the home of his daughter Margery Dodge. Tragedy marred his life again when she died in 1977. The affection which Horace, Margery's husband, and their children always had for Harry grew only stronger after her passing. Never in my experience have I seen such devotion and admiration of a father-in-law by his son-in-law.

Harry continued his active interest in and thoughtful advice to the MAA almost to the day of his death. On January 8, 1981, Horace Dodge drove Harry to the national meeting of the MAA being held at the time in San Francisco. The Board of Governors was in session when he arrived in San Francisco. It had acted on a long list of items in the morning. The meeting was considered one of the most successful ever held: the discussions were full and informative, and they led to unanimous decisions on important matters of business. By the middle of the afternoon the Board had developed a deep sense of satisfaction. Suddenly, the door opened, and Harry entered. The time and the occasion were right; President Dorothy Bernstein suspended business and introduced him as Mr. MAA amid applause from members of the Board.

Later, that evening, Harry had dinner with G. Baley Price, David P. Roselle, and Alfred B. Willcox in the Captain's Table dining room of the San Francisco Hilton. For the following account of this dinner, I am indebted to Baley Price: "The dining room was delightful; it was spacious, quiet and peaceful, and lighted and decorated in superb good taste. The meal was delicious, the conversation was pleasant; and the waiter, sensing that the occasion was something special, employed all his skills and graces to make the dinner a real party. He provided special treats from the chef, he displayed tempting desserts, and he responded with obvious pleasure to the group's every wish. After dessert, he brought, with a great flourish, a large silver bowl streaming white clouds of vapor from dry ice; the silver bowl contained four chocolate mints! The four guests enjoyed the dinner immensely; it was the end of a perfect day. Harry seemed in excellent health and spirits; when he bade the three good-night, little did they realize that they would not see him again."

HENRY L. ALDER

MATHEMATICAL MODELS: A SKETCH FOR THE PHILOSOPHY OF MATHEMATICS

SAUNDERS MAC LANE
Department of Mathematics, The University of Chicago, Chicago, IL 60637

1. Introduction. The aim of this note is to encourage a renewed study of the philosophy of mathematics, a subject dormant since about 1931. This date marks the end of a period of activity centered around what seemed to be a "crisis" in the foundations. That crisis, initiated by paradoxes such as the Russell paradox of the "set of all sets not members of themselves," led to the development of three competing schools in the philosophy of mathematics: Logicism, Formalism, and Intuitionism. In addition there were long-standing, more general philosophical doctrines: Platonism and Empiricism. Empiricism holds that mathematics is simply another branch of science, and so concludes that mathematics deals directly with the real world. However, none of these schools or doctrines in the philosophy of mathematics provides a satisfactory analysis of the nature of mathematics. Their deficiencies are eloquently analyzed in a recent article

The author received the Association's Award for Distinguished Service in 1975. See this MONTHLY, vol. 82, pp. 107-108.—*Editors*

by N. D. Goodman [4]. Other current publications in the philosophy of mathematics, from Putnam [6] to Quine [7] to Wang [10], show little new insight and no new input from mathematics.

To develop a fresh view of the philosophy of mathematics, we begin by looking at the actual state of mathematics. We do this on the grounds that a sound philosophy of mathematics ought to start with a description of what is really there. As N. D. Goodman said [4]: "Mathematics can only flourish if there is a common conception of what we are about (and only if there is an agreement that the different structures that we study are aspects of one reality)."

The various earlier philosophies of mathematics listed above each arose out of the dominant aspects of mathematics as then understood. For example, Platonism arose in Greece and applied to mathematics there because it fitted Greek geometry; it has been popular among mathematicians recently because it fitted well with the view that mathematics derives from axioms for sets. Logicism arose together with the discovery and formalization of mathematical logic. Intuitionism was the child of emphasis on numbers as the starting point of mathematics and on intuition as a basis of topology. Formalism arose with the development of axiomatic methods and the discovery that proof theory might give consistency proofs for abstract mathematics. Empiricism sprang from the 19th-century view of mathematics as almost coterminal with theoretical physics; it was much influenced by Kant's dichotomy between analytic and synthetic.

Now we search for a philosophy of mathematics better attuned to the present state of the subject.

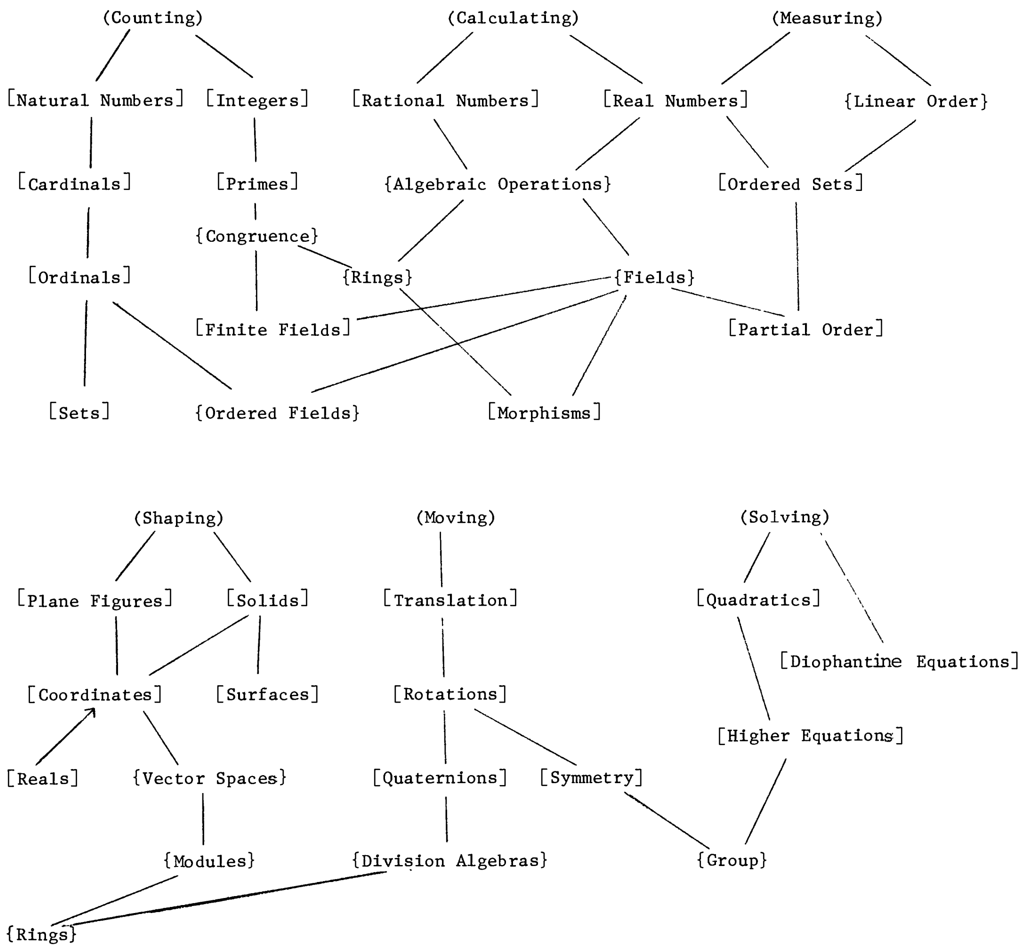
2. The Origins of Mathematics. Mathematics begins with puzzles and problems dealing with combinatoric and symbolic aspects of the general human experience. Some of these aspects turn out to be systematic and intrinsic, rather than arbitrary and tied to one context. They become the stuff of elementary mathematics. From this starting point, the subject has developed to be a deductive analysis of a large number of very different but interlocking formal structures. These structures have been derived from experience in many successive stages; by abstractions from various observations of the world, its problems, and the interconnections of these problems. These observations can be described as starting with a variety of human activities, each one of which leads more or less directly to a corresponding portion of mathematics:

Counting	:	to arithmetic and number theory;
Measuring	:	to real numbers, calculus, analysis;
Shaping	:	to geometry, topology;
Forming (as in architecture)	:	to symmetry, group theory;
Estimating	:	to probability, measure theory, statistics;
Moving	:	to mechanics, calculus, dynamics;
Calculating	:	to algebra, numerical analysis;
Proving	:	to logic;
Puzzling	:	to combinatorics, number theory;
Grouping	:	to set theory, combinatorics.

These various human activities are by no means completely separate; indeed, they interact with each other in complex ways. Table 1 gives only a partial view of this complexity, to indicate how various human activities lead to the concepts now present in algebra. The two parts of this table should (and do) fit together by many crosslinks.

On the basis of many more elaborate tables such as this, giving the origin and interconnection of mathematical ideas, we conclude that mathematics started from various human activities which suggest objects and operations (addition, multiplication, comparison of size) and thus lead to concepts (prime number, transformation) which are then embedded in formal axiomatic systems (Peano arithmetic, Euclidean geometry, the real number system, field theory, etc.). These systems turn out to codify deeper and nonobvious properties of the various originating human activities.

TABLE 1
Origins of Concepts of Abstract Algebra



For example, the notion of a group, though axiomatically very simple, reveals common properties of motion (rotation and translation groups), of symmetry (crystal groups), and of algebraic manipulations (Galois groups, Lie groups for differential equations). Many other mathematical concepts (function, partial order) are similarly both simple in structure and pervasive in application. The simplicity and the applicability are made effective by the formal treatment of the notions involved.

In this view, mathematics is formal, but not simply "formalistic"—since the forms studied in mathematics are derived from human activities and used to understand those activities.

The actual structure of mathematical ideas is an incredibly elaborate development of this simple description. Consider just the case of algebra. Algebra first involved manipulation to solve equations. Then geometry was reduced to coordinates, and thus geometrical problems to algebraic ones. Next, simple Euclidean spaces are described by vectors in two, three, and then higher dimensions. The resulting notion of a vector space, often one equipped with a (quadratic) inner product, worked even in infinite dimensions and then served to codify some of the methods of solving functional equations. The linear transformations acting on these vector spaces could be represented by matrices, which also cropped up in group theory, in numerical analysis, and in number theory. Presently vector spaces over a field were subsumed under modules over a ring.

Such spaces and modules were needed to measure connectivity phenomena in topology. These topological concepts were then borrowed again by algebraists to become homological algebra and to settle questions in ring theory and in number theory—or at least in the higher reaches of class field theory. Groups represented by matrices had extensive applications in physics, while finite groups had an elaborate structure all their own, reflecting both geometry and number theory. Additional number systems like the complex numbers or the quaternions were impossible, in view of topological arguments—but the p-adic numbers arose from a marriage of algebraic functions and algebraic numbers. This is but a small sample of the extraordinary way in which the various ideas of mathematics interlock.

Because of this elaborate interlocking pattern of ideas, each mathematical notion is tied to its empirical origins in multiple ways. As a result, no simplistic description of mathematics is adequate.

At each stage of development in mathematics, the structure at issue can be recorded as a formal deductive system. Such a system starts with axioms on a suitable list of undefined terms; in principle, it uses explicit rules of a specified logical system in order to deduce theorems and other conclusions from these axioms. Such an emphasis on the axiomatic method was not always present in mathematics. Moreover, it may seem more natural for some parts of mathematics than for others. Nevertheless, it is now always available for any part of mathematics. Our description of such systems is intended to cover use of intuitionistic or finitistic logic as alternatives to the more classical propositional and predicate calculus.

At this point, we can make a first summary of our position. Mathematics starts from a variety of human activities, disentangles from them a number of notions which are generic and not arbitrary, then formalizes these notions and their manifold interrelations. Thus, in the narrow sense, mathematics studies formal structures by deductive methods which, because of the formal character, require a standard of precision and rigor.

3. Absolute Rigor. This use of deductive and axiomatic methods focuses attention on an extraordinary accomplishment of fundamental interest: the formulation of an exact notion of *absolute rigor*. Such a notion rests on an explicit formulation of the rules of logic and their consequential and meticulous use in deriving from the axioms at issue all subsequent properties, as strictly formulated in theorems. Moreover, each derivation can be tested and understood in its own terms, independent of any reference to examples of the activity or the reality for which the axioms were designed (even though in fact that reality is usually present and often vital in suggesting how the deduction might be made). This formal character of mathematics may serve to distinguish it from all other types of science. Once the axioms and the rules are fully formulated, everything else is built up from them, without recourse to the outside world, or to intuition, or to experiment. Examination of texts of theoretical physics, biology, or other sciences clearly indicates a real difference in this regard. Such texts do not hesitate to appeal at any time to experience or intuition, while a mathematical proof stands or falls on its own, without outside reference.

An absolutely rigorous proof is rarely given explicitly. Most verbal or written mathematical proofs are simply sketches which give enough detail to indicate how a full rigorous proof might be constructed. Such sketches thus serve to convey conviction—either the conviction that the result is correct or the conviction that a rigorous proof could be constructed. Because of the conviction that comes from sketchy proofs, many mathematicians think that mathematics does not need the notion of absolute rigor and that real understanding is not achieved by rigor.

Nevertheless, I claim that the notion of absolute rigor is present. Approximation to rigorous proofs occur in many cases, in particular in the traditional proofs of Euclidean geometry. There each statement in the proof is supported by a corresponding reason or by a reference to a previous theorem. These traditional proofs failed to reach the ideal of rigor, notably at those places where a full proof would have used the axiom of Pasch to show that a desired point of intersection is really there. Nevertheless, these proofs in geometry did provide a clear model of rigor. This model was

subsequently refined in Hilbert's *Foundations of Geometry*. Both Frege, in his *Grundgesetze der Arithmetik* [3], and Whitehead and Russell, in *Principia Mathematica* [**11**], give long and careful exhibits of essentially complete and rigorous proofs. In *Principia* the idea of a rule of inference is not clearly distinguished from that of a formal axiom, but this distinction can be readily adjoined. These explicit examples were cumbersome and tedious; they clearly showed that absolute rigor was so detailed that it was a distraction and not a help to mathematical research. Nevertheless, they made the notion of absolute rigor tangible. It is the notion clearly employed in proof theory—for example, in a wide variety of completeness and incompleteness results.

The understanding of this notion of absolute rigor has in considerable measure led to the philosophical standpoints of logicism and formalism. In this way these two standpoints represent an important aspect of mathematical reality. However, these standpoints are one-sided. We emphasize that the choice of axioms, and the determination of directions in which they are to be developed, is in no wise determined by the formal structure, but rather by aspects of the world under study or by portions of the mathematician's insight or fancy.

This raises an important metaphysical issue: How does it happen that some important facets of the real world can in fact be accurately analyzed by austere deductions from axioms? In other words, how does it happen that logic fits the world; how can one account for the extraordinary and unexpected effectiveness of formal mathematics?

This issue can also be stated for particular cases:

- How is it that the formal calculation by Newtonian mechanics of the motions of bodies turns out to fit their actual motions?
- Why is it that the formal deduction of the possible groups of symmetry is matched by those groups as they occur in the world?
- Why is it that the theoretical properties of boundary value problems for differential equations describe so well so many aspects of electricity, optics, mechanics, hydrodynamics, and electrodynamics?
- How is it that the differential calculus seems to work both for physics and for the economists' problems of local maxima?

Such questions of the relations of formal logical deductions to actual events raise metaphysical problems to which I have no adequate answer.

In the practice of mathematics this notion of absolute rigor is certainly present; but a mathematician, in addition to being guided by his concepts of precision, is guided also by his understanding of the breadth and depth of his subject. By "breadth" I refer to the other objects within or without mathematics to which this subject applies, while the issue of "depth" involves judgment as to the choice of abstractions which will lead to the heart of the problems at issue. We can today clearly understand notions of rigor and formulate them in metamathematical terms, but there is no comparable analysis of breadth or depth of mathematical research. For example, why are the simple axioms for group theory so powerful?

One aspect of such an analysis is the choice of the direction for mathematical research: What topic should be studied next? On this there can be sharp opinions, for example, with Bourbaki. In the hands of Dieudonné this doctrine of chosen directions has become firm, not to say frozen. It reads: "To see whether you are doing good mathematics, look to see what Bourbaki notices; if your subject has not come to their favorable notice, it is not worth doing," Such a dogma can be stifling.

4. Multiple Models of Reality. Our thesis as to the nature of mathematics might be formulated thus: Mathematics deals with the construction of a variety of formal models of aspects of the

world and of human experience. On the one hand, this means that mathematics is not a direct theory of some underlying platonic reality, but rather an indirect theory of formal aspects of the world (or of reality, if there is such). On the other hand, our thesis emphasizes that mathematics involves a considerable variety of models. The same experience can be modeled mathematically in more than one way.

Such variant models are well known for some basic constructions. The ordinal numbers can be constructed as equivalence classes of well-ordered sets or, following von Neumann, as certain explicit sets, with 0 taken to be the empty set and each positive ordinal the set of all smaller ordinals. Because of this alternative, there is no *unique* set-theoretic description of the ordinals. However, with either description we get ordinals with the *same* behavior.

There is a similar variety in the construction of real numbers from rationals. A real number, for Dedekind, is a cut in the rationals. For Cantor or Meray, it is an equivalence class of sets of rational numbers. For Weierstrass, it is an equivalence class of sets of rationals with a bounded sum. With any of these three constructions one obtains a complete ordered field of real numbers, different constructions yielding isomorphic fields. There is no *unique* set-theoretic model for the reals.

In these cases mathematical models are determined "up to a canonical isomorphism." Indeed, that is all that matters. More generally, many mathematical constructions can be analyzed as the construction of an object which is "universal" relative to some property (i.e., which is the value of the left adjoint to a suitable functor). By its very definition, a universal object is determined *only* up to canonical isomorphism. Thus, for example, the tensor product $V \otimes W$ of two vector spaces V and W can be exhibited by an explicit set-theoretic construction—but the construction does not matter and can be immediately forgotten once the result is proved universal. All that matters is the universal property: that any bilinear map on V and W can be written in a unique way as a linear map on the tensor product $V \otimes W$.

So far, these are cases of models which are determined "up to isomorphism" or often "up to canonical isomorphism." For many mathematical purposes though, mathematicians use axiomatic systems which have many nonisomorphic models. Thus, for group theory, the depth and power of the group axioms lie in part in the fact that these axioms have many nonisomorphic models.

5. Foundations. In our view, the philosophy of mathematics is directed more at the understanding of the nature of mathematics than at a "foundation" of mathematics. Nevertheless, our emphasis on the fact that finished mathematics is formal is close to questions about foundations. The clear understanding of formalism in mathematics has led to a rather fixed dogmatic position which reads: Mathematics is what can be done within axiomatic set theory using classical predicate logic. I call this doctrine the *Grand Set Theoretic Foundation*.

In a preliminary version, this arose in the 19th century with Weierstrass, Dedekind, and Frege: Start with finite cardinal numbers, perhaps defined by set theory, and build up from them the natural numbers, then the integers, then the rational numbers as pairs of integers, and then the real numbers, say as Dedekind cuts in the rationals. This careful construction of the real numbers was long accepted as standard in graduate education in mathematics, even though many mathematicians did not much believe in it. They also were not always aware that this construction did not get the real numbers from natural numbers alone but had to use set theory extensively on the way. While paying lip-service to this real-number foundation, it was felt that a real number is really a point in the preexisting geometric continuum and not just a formal Dedekind cut in the rationals. This viewpoint can be expressed intuitively (as a geometric insight) or formally: Do not construct the reals, but describe them axiomatically as an ordered field, complete in the sense that every bounded set has a least upper bound.

The next step in the grand set-theoretic foundation included sets (or classes) with logic and was initiated in the work of Frege. This direction reached a crisis with the 1900 discovery of the Russell paradox of the set of all those sets not members of themselves. Actually, this paradox itself

was settled very soon, in 1908, independently and in two different ways, by Russell's publication of his *Theory of types* [8] and by Zermelo's publication, also in 1908, of his *Axioms of set theory*.

It took a considerable period before this solution and system was shaken down and well formulated through the work of Skolem, Fraenkel, Paul Bernays, and others. Even in the 1940's, with the growth of abstract algebra, axiomatic set theory was not regarded as a central doctrine. It was not until about 1950 that the *Grand Set Theoretic Foundation* was finally complete and officially accepted under the slogan which might have read: "Mathematics is exactly that subject which can be developed by logical rules of proof from the Zermelo-Fraenkel axioms for set theory." This foundation scheme had its popular version in the "new math" for schools. It also had its philosophical doctrine, a version of Platonism, that the world of sets is that constructed in the standard cumulative hierarchy of all ranked sets. Here one begins with an understanding of the empty set and the ordinals and uses the power set (the set Px of all subsets of x) to construct the iterated power sets of the empty set $\varnothing$:

$$R_0 = \varnothing, R_1 = P(\varnothing), R_2 = P(P\varnothing), \ldots, R_n = P^n\varnothing, \ldots, R_\omega = \bigcup_{n \in \omega} R_n,$$

and so on through all the ordinal numbers $0, 1, 2, \ldots, 10, \ldots,$ using, at each limit ordinal, the union of all preceding sets. Thus for each ordinal α one has the collection R_α of sets of rank α. In this way, the hierarchy of sets is presented as a cumulative type theory, with R_α as all sets of type α or *less*. The Zermelo-Fraenkel axioms are then (a selection of) the facts true for all sets in this hierarchy. This is sometimes claimed to describe the ultimate Platonic reality which underlies all mathematics: Perhaps the Zermelo-Fraenkel axioms do not describe everything, but with a little more insight we will understand all the axioms necessary and then at least in principle all mathematical problems can be settled from the axioms.

It is my contention that this *Grand Set Theoretic Foundation* is a mistakenly one-sided view of mathematics and also that its precursor doctrine (Dedekind cuts) was also one-sided. This grand formulation does succeed in recording a view of mathematical rigor, but by emphasizing this it misses other important points about the nature of mathematics.

We may list various difficulties with the grand foundation as follows:

First, it does not adequately describe which are the relevant mathematical structures to be built up from the starting point of set theory. *A priori* from set theory there could be very many such structures, but in fact there are a few which are dominant, one list being provided by Bourbaki's "mother structures." Some mathematical structures (natural numbers, rational numbers, real numbers, Euclidean geometry) are intended to be unique but other structures are built to have many different models: group, ring, order and partial order, linear space and module, topological space, measure space. The "Grand Foundation" does not provide any way in which to explain the choice of these concepts (such a choice depends on the "breadth" parameters and the relation to the outside world). The grand foundation also does not recognize the common notions which appear in different types of structures, as for example the sense in which "universal" constructions (that is, adjoint functors) appear in many different places: in the construction of a vector space on a given set as basis, of a free group on a given set as generators, of the Stone-Čech compactification of a given space and of the tensor product of two vector spaces. Further development of these ideas in a positive direction seems to require the provision of a big table of mathematical structures. This table would give not only the interrelations and commonalities between the structures, but also their origins in activities arising in the world. This seems more hopeful than trying to find a formal description which will designate the mathematical structures to be studied.

Second, set theory is largely irrelevant to the practice of most mathematics. Most professional mathematicians never have occasion to use the Zermelo-Fraenkel axioms, while others do not even know them. If they did know the axioms, they would rapidly observe that most of the mathematics they do could be satisfactorily based on a much weaker set of axioms—say the

Zermelo system in which the replacement axiom of Fraenkel is dropped in favor of the weaker comprehension axiom. The comprehension axiom does allow the frequently used formation of the set of all x with a specified property $P(x)$, where with Fraenkel and Skolem a property is anything expressed by an ε-formula of the first-order predicate calculus. However, for mathematical purposes it suffices in most cases to use only these formulas where the quantifiers are bounded (i.e., where $\forall x$ or $\exists x$ is applied only for x in a given set).

Thus, technically, there is not one preferred system of axioms for the set theory used by mathematicians. This, however, is not the real point—which is that in practice set theory is not the grounds of all mathematics, but of just one highly specialized branch of mathematics.

The *Grand Set Theoretic Foundation* of mathematics has other, more technical disadvantages. It does not answer the difficulties presented by the Gödel incompleteness theorem. It is not strong enough to take into account some of the large constructions on the fringe of mathematics. For example, one would like to form the category of all sets (essentially the set of all sets and of all functions between sets). This can be done by speaking of the "class" of all sets. This device, however, will not yield bigger constructions, such as the category of all categories. That can be managed by a different device: Assume that there is a (Grothendieck) universe containing all (ordinary) sets, build bigger sets out of this universe, and then form the category of all categories contained in this universe. These devices to make set theory include the fringes seem artificial.

The set-theoretic approach is by no means the only possible foundation for mathematics. Another approach is to formulate axioms not on set membership, but on the composition of functions. This results in an axiomatization of the category of all sets. The resulting axioms (those for an "elementary topos") describe cartesian products, power sets, and the like, by certain "universal" properties. For this reason they probably give better insight into the conceptual form of mathematics than does set theory. There may well be other possible systematic foundations different from set-theoretic or categorical ones.

The final difficulty with the Grand Foundation is that it does not account for what E. P. Wigner has termed the unreasonable success of mathematics in its applications.

6. Cantorian Set Theory. Many students of set theory do not follow what I have called the "Grand Set Theoretic Foundation" but instead follow Cantor to emphasize the intuitive notion of a set as a collection which is a real object in its own right. For them set theory is not subsumed by the Zermelo-Fraenkel axiom system or by any other first-order formal system. It may be studied formally by other means; using infinitary languages or second-order logic. Such Cantorian sets are just as real as numbers. Indeed, one might say that number theory is formalized only in part by Peano's arithmetic in just the way set theory is formalized, but only in part, by Zermelo-Fraenkel. (There can be true properties of whole numbers not demonstrated in Peano arithmetic.)

From this point of view, set theory is just another branch of mathematics. If in this view set theory is not taken to be *the* foundation of mathematics, it can be assimilated with our proposal that mathematics consists of formal disciplines derived from a variety of human activities. Here the relevant activity is that of "collecting" things into "totalities."

However, this Cantorian point of view is often taken to concern a Platonic world of sets. This does not fit our proposal.

7. Multiple Models for Set Theory. By now there are substantially different models of set theory, satisfying one or another special axiom— the axiom of constructibility, Martin's axiom, or the axiom of determinateness. The striking result of these technical developments is that different models of set theory give different answers to specific mathematical problems. The continuum hypothesis is true on the Gödel axiom of constructibility, but false in certain Cohen models of set theory. Whitehead's problem provides another striking example. He considered a homomorphism $f: A \to G$ of one abelian group A onto another such group G, in the case when the kernel is just the (additive) group of integers. In case G is a free abelian group, the epimorphism splits (that is, there

is a homomorphism $h: G \to A$ with $fh = 1$). Whitehead asked: Conversely, if such an f always splits, is G free? It now turns out that the answer may be yes or no, depending on the model of set theory. This is one of many striking cases where explicit mathematical problems have different answers, depending on the model used for set theory. (See Eklof [**2**].)

Mathematics, we hold, deals with multiple models of the world. It is not subsumed in any one big model or by any one grand system of axioms.

The idea that set theory is relative is not new; it was clearly stated for axiomatic set theory by Skolem in 1922 [**9**]. We intend simply to draw some of the philosophic consequences of that relativity. For the Platonist, there is a real world of sets, existing forever, described only approximately by the Zermelo-Fraenkel axioms or by their modifications. It may be that some final insight will give a definite axiom system, but the sets themselves are the underlying mathematical reality.

In our view, such a Platonic world is speculative. It cannot be clearly explained as a matter of fact (ontologically) or as an object of human knowledge (epistemologically). Moreover, such ideal worlds rapidly become too elaborate; they must display not only the sets but all the other separate structures which mathematicians have described or will discover. The real nature of these structures does not lie in their often artificial construction from set theory, but in their relation to simple mathematical ideas or to basic human activities.

Hence, we hold that mathematics is not the study of intangible Platonic worlds, but of tangible formal systems which have arisen from real human activities.

8. Models of Geometry. Space provides a striking example of the multiple variety of mathematical models. The original human experience of space is vague and varied: Space is extensive and hollow, both fixed and the locus of movement. With Euclidean geometry it is analyzed axiomatically as a receptacle: Space is described in terms of the things (lines, triangles, circles) which can be pushed around within it. With non-Euclidean geometry came the possibility of a different deductive model of the "same" space. In a different direction, the description of the plane and the three-space by means of Cartesian coordinates led to an analysis of much more general figures within space: those given by general algebraic equations or by other functions, including in particular poorly behaved functions (curves without tangents). This analytic approach also presently indicated that those original geometric intuitions of space also applied to space of more than three dimensions—and even to infinite dimensional spaces. Here, too, space is apprehended partially in geometric terms, and partially—by vector analysis—in algebraic fashion. Thus there are many mathematical models of space (Mac Lane [**5**]).

The case of topological spaces and manifolds is especially striking. First came the general notion of a metric space, motivated, it seems, by the use of a metric for function spaces for the calculus of variations and for integral equations. Then came a striking discovery: The continuity of a real valued function f on a metric space M can be defined wholly in terms of the open subsets of M. It was this discovery, combined with the study of Riemann surfaces, which led to the definition of topological spaces. This definition represented considerable extension of the notion of geometry.

However, for other parts of geometry one needed algebraic functions or differentiable functions on the space—and these classes of functions (apparently) cannot be described just in terms of their action on subsets of that space. One must instead specify for each open subset of the space all the good functions (differentiable, analytic, or algebraic, as the case may be) on that subset. These specifications amount to defining a sheaf on the space. In this way, a differentiable manifold can be described as pieces of Euclidean space, pasted together so that the appropriate sheaves match. Similarly for algebraic geometry a "scheme" is described by pasting together suitable affine pieces so that the sheaves match. Thus the intuitive idea of a "space" for differential or algebraic geometry can be adequately formalized only by sophisticated and deep notions, such as those of sheaf theory.

9. Breadth, Clarity, and Depth. Let me return to the philosophical issues. We hold that logicism, formalism, and Platonism have been too much dominated by the notions of set theory and deductive rigor. A balanced philosophy of mathematics should complement these ideas with others. The others we tended to list as three: breadth, clarity, and depth.

All three become important because of the extent of abstraction in mathematics. Abstraction consists in formulating essential aspects of some subject matter in terms of suitable axioms. Such abstraction can take place in several successive stages, interlocking different branches of mathematics. However, to be well directed or relevant that abstraction needs these three qualities.

The breadth of a mathematical notion refers to the variety of the situations in which it is to apply and to the pertinence and relevance of the abstraction made. It carries also the caution that deductions of theorems are guided not just by rigor but by the intent of the applications or by the origin of the abstraction.

Second, abstraction has increased the need for clarity in presentation; if the object of study is abstract, it must be understood not by its intuitive content but by its precise and abstract description.

Clarity goes beyond the precision of rigor to a clear ordering of ideas. The development of abstract mathematics, especially after 1920, is in this view a reflection of the necessity for such clarity. When geometry was the geometry of a three-dimensional real world, there could be continual appeals to the real world. Now, done in greater generality, it must be done rigorously and exactly; this means also that it must be clear and perspicuous.

The depth of a mathematical notion refers to the way in which that notion gets at the nonobvious, more fundamental structures and concepts underlying the problems at issue—as group theory underlies symmetry or as uniform continuity is subtly involved in many questions of real analysis. The study of manifolds in differential geometry and in algebraic geometry offers another example of the discovery of deeper notions. Initially one thinks of a manifold as a suitable smooth set of points spread out in some given ambient space. Later on one forgets the ambient space and considers the manifold in terms of the well-behaved functions which can be defined on it, as well as the germs of such functions at each point. This study in turn leads to the sheaf of germs of well-behaved functions on the manifold, and so to the deeper ideas of sheaf theory.

The depth of a mathematical notion may well change with time. For example, in the late nineteenth century the notion of uniform continuity seemed hard. It now seems easier, and is often dismissed as being simply a change in the position of an existential quantifier.

On the basis of this observation, we attempt a definition of mathematics about as follows:

Mathematics consists in the discovery of successive stages of the formal structures underlying the world and human activities in that world, with emphasis on those structures of broad applicability and those reflecting deeper aspects of the world.

In detail, mathematical development uses experience and intuitive insights to discover appropriate formal structures, to make deductive analyses of these structures, and to establish formal interconnections between them. In other words, mathematics studies interlocking structures. Because of the depth and of the distance from immediate concerns, mathematical treatments need be not only rigorous but also endowed with conceptual clarity.

References

1. F. Browder and S. Mac Lane, The relevance of mathematics, in Mathematics Today, ed. by Lynn Steen, Springer-Verlag, 1978, pp. 323–349.

2. Paul C. Eklof, Whitehead's problem is undecidable, this MONTHLY, 83 (1976) 775–787.

3. G. Frege, Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, vol. 1, Jena, 1893; vol. 2, Jena, 1903.

4. N. D. Goodman, Mathematics as an objective science, this MONTHLY, 86 (1979) 540–555.

5. S. Mac Lane, Mathematical models of space, American Scientist, 68 (1980) 184–191.

6. Hilary Putnam, Philosophy of mathematics, in Current Research in the Philosophy of Science, ed. by Peter Asquith and Henry E. Kyburg, Jr., Philosophy of Science Association, East Lansing, Michigan, 1979, pp. 386–398.

7. W. V. Quine, Ontological Relativity and Other Essays, Columbia University Press, New York, 1969.

8. Bertrand Russell, Mathematical logic as based on the theory of types, Amer. J. Math., 30 (1908) 222-262.
9. Th. Skolem, Einige Bemerkungen zur axiomatische Begründung der Mengenlehre. Fifth Congress of Scandinavian Mathematicians, 1922, Helsingfors, 1923, pp. 217-232.
10. Hao Wang, From Mathematics to Philosophy, London, Routledge and Kegan Paul, 1971.
11. A. N. Whitehead and Bertrand Russell, Principia Mathematica vol. 1, Cambridge University Press, Cambridge, England, 1910; 2nd edition, 1925.
12. E. Zermelo, Untersuchungen über die Grundlagen der Mengenlehre I, Math Ann., 65 (1908) 261-281.

COMPUTER SCIENCE, MATHEMATICS, AND THE UNDERGRADUATE CURRICULA IN BOTH

ANTHONY RALSTON
Department of Computer Science, SUNY at Buffalo, Amherst, NY 14226

There is a simple and basic fact about a computer which will, in the decades and centuries to come, affect not so much what is known in mathematics as what is thought important in it. This is its finiteness.

Wallace Givens [**7** (1966)]

We have thought of the calculus as the beginning, the gateway. But should it be?

H. O. Pollak [**21** (1978)]

If new disciplines may be described as emerging from old ones, then computer science may be said to have sprung mainly from mathematics, although, of course, the influence of electrical engineering was also considerable. In recent years, however, the ties between mathematics and computer science have been steadily weakening. This has been coupled with a declining belief on the part of some (most?) computer scientists in the importance of mathematical training and mathematical tools in the education of computer scientists and in research in computer science. Some of the evidence for this "declining belief" will be considered later in this paper. Here I note only my belief that, contrariwise, the importance of mathematics in computer science is and should be growing rapidly. It is this belief which motivates much of this paper.

While thus far in the three decades of computing and in the two of computer science there has been considerable discussion of the influence, or lack of it, of mathematics on computer science, there has been little discussion of what should—or should not—be the influence of the development of computer science on mathematics research and education. The other major motivation of this paper stems from my belief that the growth of computer science should be having, but has not had, a profound effect on undergraduate mathematics education.

1. The Role of Mathematics in Computer Science Education. In the 1960's, in addition to the emerging departments of computer science themselves, there were numerous examples of departments of statistics and computer science, or applied mathematics and computer science, or computer programs and options within departments of mathematics. And this was in contrast to a

Anthony Ralston received his Ph.D. in Mathematics under Eric Reissner at MIT. He has held positions at Bell Laboratories and Stevens Institute of Technology and is now Professor of Computer Science at SUNY at Buffalo. He has been a visiting faculty member at the University of Leeds, the Institute of Computer Science of the University of London, and University College, London. He is a former president of the Association for Computing Machinery and of the American Federation of Information Processing Societies. Most of Professor Ralston's early research papers were in numerical analysis. In recent years, after publishing mainly on strictly computer science topics, he has "returned" to mathematics with current interests in discrete mathematics and mathematics education. He is the author or editor of eight books, including *A First Course in Numerical Analysis* (second edition, with P. Rabinowitz, McGraw-Hill, 1978) and the *Encyclopedia of Computer Science* (Van Nostrand Reinhold, 1976).

few—but very few—joint efforts with engineering departments. In 1980, however, we find that, while some smaller colleges and universities, most of which have never had the resources to start separate departments of computer science, still house computer science within a department of mathematics, there is no example in the United States of a computer science program with a significant research reputation either housed in a department of mathematics or jointly with any mathematics-like department. But, in addition to stand-alone departments of computer science, which are the general rule, there are several significant examples, MIT and Berkeley being the best known, of joint departments of computer science and electrical engineering. (These are always called departments of electrical engineering and computer science, but that is only because computer scientists have made relatively little headway in teaching engineers the meaning of lexical order.)

Why have the ties between mathematics departments and computer science departments weakened so much over the past decade? For a variety of reasons, I think:

1. There has been a relative decrease in importance in computer science of some of its more traditional mathematical areas, such as numerical analysis.
2. The dominant role played by mathematicians in the establishment of computer science as a discipline explains the early close association of the two disciplines. But the difficulty computer science had in being recognized as a separate discipline (and not just as a branch of mathematics) at many universities, and in the academic and scholarly worlds more generally, explains the desire of many computer scientists to emphasize the distinctness of computer science and, therefore, to form separate departments.
3. Beginning with the early 1970's computer science faculties became increasingly populated by those whose degrees were in computer science. Data on this are lacking until the mid-seventies but by then [**26**] almost 40 percent of the faculty in Ph.D.-granting departments of computer science held Ph.D.'s in computer science. Moreover those with Ph.D.'s in mathematics consisted of only about 20 percent of the total faculty, and this percentage appeared to be dropping rather rapidly. Certainly by the mid-seventies there were very few computer science faculty members with *recent* mathematics Ph.D.'s. Thus the personnel ties between mathematics and computer science have been steadily attenuated since the establishment of computer science as a discipline.
4. Finally, and not least, mathematics departments have not been, and are not now, very hospitable to the ideas and techniques of computer science, surely much less so than many electrical engineering departments. This was particularly true in the sixties, as mathematics departments focused increasingly on abstract branches of mathematics. Although the seventies saw some redress of this trend, there is little doubt that computer scientists, the vast majority of whom are pragmatically and applications oriented, to a considerable degree still see little in the prevailing philosophy of departments of mathematics that implies an understanding of, or a sympathy for, the interests and aims of computer science. This is a generalization, of course, but a correct enough one to explain much of the current lack of close collaboration between mathematics departments and computer science departments.

The loosening of the ties or, if you will, the cooling of the relationship between mathematics and computer science as disciplines has been reflected in the place of mathematics in the computer science curriculum (to the detriment, I believe, of both disciplines). In the first important publication on the undergraduate curriculum in computer science [**1**], there was a clear statement of the distinction which computer scientists saw between mathematics and computer science, a distinction which reflects the trends in the two disciplines in the sixties:

> The mathematician is interested in discovering the syntactic relation between elements based on a set of axioms which may have no physical reality. The computer scientist is interested in discovering the pragmatic means by which information can be transformed to model and analyze the information transformations in the real world.

Nevertheless, because with little or no exception the authors of this report had had their academic

training in mathematics, they believed strongly in the importance of mathematics in computer science education. This is reflected in their final report known as Curriculum 68 [**2**].

It is generally believed and certainly many people have testified, both formally and informally, that Curriculum 68 had far-reaching effects on the development of undergraduate computer science curricula in the United States and, indeed, elsewhere. The attitude of the Curriculum 68 committee toward mathematics is clear:

> The Committee feels that an academic program in computer science must be well based in mathematics since computer science draws so heavily upon mathematical ideas and methods. The recommendations for required mathematics courses given below should be regarded as minimal; obviously additional course work in mathematics would be essential for students specializing in numerical applications.

The "minimal" mathematics requirement in Curriculum 68 consists of six required courses—three of the four semesters of the calculus sequence, a course in probability, a course in numerical calculus and one in discrete structures—plus two of four other courses—the fourth semester of the calculus sequence, a semester of advanced calculus, a course in algebraic structures, and one on probability and statistics. Two comments are pertinent here:

1. All the courses, with the exception of discrete structures and numerical calculus, were and are standard mathematics (or statistics) department courses. Indeed, these eight courses are specified to be precisely those recommended by CUPM [**15**].
2. The discrete structures course is the only recognition in Curriculum 68 that the mathematics needs of computer science are different from those in the physical sciences. However, the description of this course in Curriculum 68 read: "Review of set algebra including mappings and relations. Algebraic structures including semigroups and groups. Elements of the theory of directed and undirected graphs. Boolean algebra and propositional logic. Applications of these structures to various areas of computer science." This illustrates that, while there are distinctions between it and the usual subject matter of abstract algebra courses (e.g., graph theory), the course as a whole was not unlike many abstract algebra courses. The attitude toward mathematics of the developers of Curriculum 68 was quite appropriate to the state of computer science in 1968. This state can be quickly summarized as follows: Mathematics education is central to the undergraduate program in computer science; the mathematics to be taught to undergraduate computer scientists should be essentially that traditionally taught to physical science majors.

During the past decade there have been two somewhat contradictory trends in the relationship between computer science and mathematics. On the one hand, the place of mathematics in the undergraduate computer science curriculum was steadily weakened, the culmination of this being Curriculum 78, which we consider below. On the other hand, there has been an increasing recognition of the relative importance of discrete mathematics to the computer scientist. For example, in his 1974 paper Knuth [**13**] notes: "Discrete mathematics, especially combinatorial theory, has been given an added boost by the rise of computer science." And he goes on to say, "I have been naturally wondering whether or not the traditional curriculum (the calculus courses, etc.) should be revised in order to include more of these discrete mathematical manipulations or whether computer science is exceptional in its frequent application of them."

The latter part of the 1970's saw the publication of two major undergraduate curriculum studies in computer science, the first (in 1977) a report of the IEEE Computer Society [**10**] and the second the long-awaited revision of Curriculum 68, called Curriculum 78—although finally published in 1979 [**3**]. Noteworthy about the former is its almost total lack of discussion of the mathematics requirements for a program in computer science and engineering. Although portions of a discrete structures course and an analysis of algorithms course are part of the core curriculum, these are part of a theory of computation sequence. Moreover, in the only sample program given in the report, it is assumed that students will take the usual calculus–linear algebra sequence in the first two years. And there is no explicit probability or statistics requirement.

Curriculum 78, while it contains considerably more discussion of the mathematics requirements for an undergraduate program than the IEEE Computer Society curriculum, is nevertheless quite disappointing relative to Curriculum 68. A comparison between Curriculum 68 and Curriculum 78 discloses that [23]:

1. Whereas Curriculum 68 required the student to take eight semester-length mathematics courses, Curriculum 78 requires only five mathematics courses.
2. Whereas the mathematics courses in Curriculum 68 formed an integral part of prerequisite structure, in Curriculum 78 there is no mathematical prerequisite for any undergraduate computer science course, with the exception of three advanced and clearly quite mathematical courses.
3. Curriculum 78 continues the reliance of Curriculum 68 on continuous mathematics.

We have, then, the situation that, while many computer scientists believe that more, not less, mathematics is needed in undergraduate computer science education and that the traditional calculus sequence is perhaps not the most appropriate mathematics for this curriculum, the main curriculum committees of the professional societies in computing are either ignoring this question or taking an almost opposite point of view.

A central thesis of this paper is that the role of mathematics in computer science needs to be both strengthened and redirected. The argument for redirection is made later in this paper. But the importance of redirecting the curriculum must be proportional to the importance of mathematics in the computer science curriculum itself. If mathematics is just not very important in the undergraduate computer science curriculum, then it would be hard to be very excited about changing this component of the curriculum.

How important then is mathematics in the undergraduate computer science curriculum? Curriculum 78 will not only bolster the view of many outside the discipline that Computer Science = Programming but it will also surely strengthen the view that, while mathematics may be good for the soul (and maybe for the mind also), it has little direct relevance to undergraduate computer science. My view on this can best be expressed by the following quotation [23]:

> The principles and theories of any science give it structure and make it systematic. They should set the shape of the curriculum for that science, for
> —only in that way can they provide a framework for the mastery of facts, and
> —only in that way will they become the tools of the practicing scientist.
> This is as true for computer science as it is for mathematics, for the physical sciences, and for any engineering curriculum. Inevitably, for any science or any engineering discipline, the fundamental principles and theories can only be understood through the medium of mathematics.

But perhaps the strongest case for a strong mathematical component in undergraduate computer science can be made by noting that, even if the false view that Computer Science = Programming were true, the case for more mathematics in computer science would be overwhelming (see, for example, [6]). The single biggest challenge faced in the overall area of programming is how to provide it with some theoretical (i.e., mathematical) underpinnings that will enable software to be developed and maintained more efficiently and with a higher level of reliability than is normally attained nowadays. Many of the most important research areas in computer science, such as the work in program verification and structured programming, are concerned with these problems. And all of this is highly mathematical. Moreover, the need to teach computer science students how to analyze and verify algorithms is becoming increasingly clear. And this, too, requires intensive mathematical analysis. Finally, at the risk of noting the obvious, two purposes of all mathematics education are to instill in students the meaning of rigorous thinking and an appreciation of abstraction. The importance of both of these to programming—and to computer science more generally—can hardly be overstressed.

The mathematical tools and methodology required by computer science students are, with few exceptions, not the classical tools and methodology of (continuous) analysis. Rather they are the tools of discrete analysis, and it is this type of mathematics which should be the heart of the

undergraduate mathematical training of computer science students. In Section 3 we shall develop this idea at length. But first, in the next section, we shall consider the influence of computing and computer science on the mathematics curriculum.

2. Computing, Computer Science, and the Mathematics Curriculum. At the risk of some modest oversimplification, the following describes the first two years of the undergraduate mathematics curriculum during the past two-thirds of a century. (This paper is a shortened version of a much longer paper by the author [**22**] which contains an extensive review supporting these conclusions. The longer report may be obtained from the author by writing to the Department of Computer Science, 4226 Ridge Lea Road, Amherst, NY 14226.)

1. The essential change over this period has been from a college algebra–trigonometry–some calculus in the first two years situation, to a two-year calculus–analytic geometry sequence, to the more common calculus–linear algebra sequence of today.
2. This change has been accompanied by changes in approach in the presentation of material from the relatively informal and concrete, to the rigorous and abstract, and now back to a more informal approach. But, whenever and however taught, calculus has remained the entrance point to all study of higher mathematics for professional mathematicians, scientists, engineers, and, more recently, social, behavioral, and management scientists.

As far as computing is concerned, two trends over the past two decades are discernible in the attitude of mathematicians toward the effect computing might have on the mathematics curriculum. One is the attempt to incorporate computing into the traditional curriculum, most notably in the calculus sequence. This trend is exemplified by many articles published over the past 10 or so years. Gordon [**8**] has a good bibliography of articles on this subject. Another and rather more recent trend concerns the attempt to combine computing and computer science with the mathematics curriculum to provide a more broadly based major centered on mathematics.

One may argue about whether or not mathematicians have embraced the use of computers in their courses as rapidly or intensively as they should have. But our purpose here is only to note that the use of computers in the calculus and other courses, however desirable and valuable, has not had any marked effect on the subject matter of undergraduate mathematics itself. Indeed, even when the mathematical community has considered programs in "computational mathematics" [**16**], the essential subject matter for the first two years has hardly changed. The only nontraditional mathematics course in the CUPM Undergraduate Program in Computational Mathematics is a course in combinatorial computing. In fact, the mathematics requirements of this program are quite similar to those of Curriculum 68.

A number of mathematicians have recognized that the computer allows a number of areas of mathematics to be taught experimentally (e.g., McKenna [**19**]). But there has been little realization that the advent of computers *and* computer science might suggest some fundamental changes in the undergraduate curriculum itself.

It is natural and proper that mathematicians have been worried and frustrated by declining enrollments, particularly at the upper division undergraduate and graduate levels. The frustration and worry were succinctly juxtaposed in a 1979 report on the role of applications in the undergraduate curriculum [**20**]:

> Mathematics is playing an increasing role in the physical and management sciences and in engineering, in both academic and nonacademic spheres. Yet enrollments are declining drastically in the mathematics major and in many of the traditional postcalculus courses.

An even more pessimistic outlook is contained in a view toward 1984 in [**18**, p. 42] which predicts that by then the only service courses left to mathematics departments would be "remedial algebra... and two semesters of calculus."

A natural response to these unwelcome trends is to try to make the mathematics curriculum more attractive to students. One general approach, already in progress before the advent of declining enrollments, was to emphasize less abstract approaches and subject matter. Related to

this are attempts to introduce more applied mathematics and applications of mathematics into the curriculum. But by far the most popular and recommended response—not, by the way, unrelated to the other two—is to introduce more computing and computer science into the mathematics curriculum, sometimes into programs called "mathematical sciences." My own belief is that a mathematical sciences curriculum not only makes disciplinary sense but is a proper and reasonable response to enrollment difficulties and societal needs. There are, however, two pertinent comments about such curricula:

1. They imply a belief on the part of mathematicians that computer science is a "mathematical science" (whatever that is) in much the same sense as statistics is. But those portions of computer science which are growing most rapidly (i.e., computer systems and software) are, despite their great needs for mathematical underpinnings, not mathematics in any usual sense of the word. Indeed, one of the reasons for the aforementioned greater affinity of computer science departments with departments of electrical engineering than with mathematics departments is because the core areas of computer science are closer in outlook and approach to engineering than to mathematics.
2. The curricula which have been proposed for the mathematical sciences are, with the typical exception of one discrete mathematics or combinatorial computing course, quite traditional in the sense that the courses they propose are typically amalgams of traditional mathematics, computer science, and statistics courses. It would be wrong to be too critical of this, because, in times of limited resources, there are obvious attractions to new programs which require the offering of few new courses. Nevertheless it is my belief that, through resource inhibitions or lack of imagination, these programs have missed an opportunity to be far more useful and innovative than they are.

The conclusion I would emphasize here is that there is no evidence of any significant impact of computing or computer science on the essential subject matter of the undergraduate mathematics curriculum, particularly the first two years. I believe that there should be such an impact and in Section 4 I return to this subject.

3. Mathematics for Computer Science. In order to approach the question of how the undergraduate mathematics curriculum itself might be restructured, it is instructive to consider first the mathematics requirements of an undergraduate computer science curriculum.

The computer science courses taken by a student during the first two undergraduate years generally consist roughly of the following:

1. A one-semester or, increasingly commonly, a one-year sequence introduction to computer science which emphasizes
   —algorithmic processes
   —learning to program in one or more higher level languages
   —programming methodology and style
2. A one-semester course on assembly language programming and machine organization.
3. A one-semester data structures course.

What mathematics is desirable and/or necessary to support such courses? To begin, we note the total absence of any *need* for calculus in any of these courses. This is not to say that, for purposes of examples, it would not be useful to draw upon the calculus. But it is certainly not necessary, and the potential for examples in all the courses listed above is so great that an inability to use calculus is no great loss.

We have argued earlier that mathematics is more important than ever for computer science students. But if not calculus, then what? The answer follows most easily from the observation that one definition of computer science favored by some is the study of algorithms [**13**]. Whether or not you find this an acceptable definition, it is undeniable that the design, analysis, and verification of algorithms should play a crucial role in the first year of computer science and an important role in many subsequent courses. While the importance of algorithms can also be stressed in calculus courses (see, for example, Rosser [**24**]), not only is discrete mathematics rather than continuous

mathematics the main mathematical component of most computer algorithms but by far the fastest growing area of algorithmics is that which needs discrete mathematics.

Since discrete mathematics is one of those amorphous terms with various different meanings to different people, we should first indicate as clearly as possible what we shall take its meaning to be in the remainder of this paper. As used here, discrete mathematics covers those branches of mathematics which focus mainly or entirely on discrete objects such as combinatorics, graph theory, abstract algebra, linear algebra, number theory, set theory, and discrete probability. The term also naturally implies an emphasis on discrete analysis in contradistinction to the traditional emphasis on the analysis of continuous functions.

To get some idea of the kinds of mathematics needed to support the computer science courses listed above, let us list some topics which require mathematics and which are always, or often, covered in such courses:

- —analysis of algorithms, including the use of induction, discrete probability, often the summation or manipulation of series, and various aspects of basic combinatorial theory;
- —verification of algorithms, including the generation of logical assertions about algorithms and informal proofs of these;
- —numbers and number systems, particularly the discrete number system which is the basis of all computer arithmetic;
- —simple queueing theory used to discuss basic aspects of scheduling;
- —random number generation, since random number generators are often used in problems given in basic computer science courses.

In addition, various other topics in discrete mathematics may also be covered.

Most mathematicians would agree, I think, that, for students intending to major in disciplines other than mathematics, the mathematics they study during their first two undergraduate years should support and, where possible, be coordinated with corresponding courses in their intended major discipline. Therefore, in the absence of mathematics courses which support the first two years of computer science, instructors in computer science are usually forced to cover the necessary mathematics themselves. Inevitably this has two bad results:

- —the time spent on the mathematics is a distraction from the essential computer science subject matter in addition to lessening the computer science component of the courses;
- —in order to minimize the distraction and the loss of computer science subject matter, the mathematics tends to be covered cursorily, incompletely, or both.

It is hardly news that the need to shore up a student's understanding of subject B in a course on subject A is usually unsuccessful in imparting real understanding of B and detracts from the coverage of A.

The sequence of courses to be outlined briefly now is only one possible solution to this problem. This sequence is discussed in extensive detail in [**22**]. Here, then, is an outline of a possible two-year sequence in discrete mathematics.

First Year

A one-year course in discrete mathematics covering:

Algorithms and Their Analysis
Introductory Mathematical Logic
Limits and Summation
Mathematical Induction
The Discrete Number System
Basic Combinatorial Analysis
Difference Equations and Generating Functions
Discrete Probability
Graphs and Trees
Basic Recursion and Automata Theory

Remarks.

1. The "glue" which should be used (and which the author has used in such a course) to hold together an (otherwise) disparate set of topics is the notion of an algorithm and the analysis of algorithms.
2. The above should be judged by its general flavor and the philosophy it implies and not by its details. With Berztiss [**4**], I believe that the syllabus itself "is of no great importance, as long as the basic intents are being satisfied." This syllabus should only be considered as an example of one of many possible syllabuses and should be judged as to
   —whether the topics chosen are important (rather than most important)
   —whether they are accessible to first-year students
   —whether they will tend to develop mathematical maturity and sophistication as well as the calculus
   —whether they are appropriately supportive of the computer science curriculum.

Second Year

For this year we recommend a semester each of linear algebra and abstract algebra. The former needs little more said about it. It should be closely modeled on the current one-semester course so prevalent in the second-year mathematics curriculum, partly because this would involve minimal curriculum upset and partly because this course is quite relevant to the mathematical needs of computer science students.

Since abstract algebra is usually a third-year, two-semester sequence, it is not quite so easy to specify what to do here. Obviously a single semester of the usual sequence is a possible solution motivated both by the desire to avoid unnecessary curriculum upset and because, as we shall argue in the next section, the curriculum outlined here is also intended to be reasonable for mathematics majors. Nevertheless, because computer science students do have specific needs in the area of abstract algebra, it is worthwhile to outline the set of topics which such a course should include:

Functions, relations, and equivalence
Fields, rings, and ideals
Groups, including at least permutation and cyclic groups, semigroups and cosets
Lattices
Algebras generally, including Boolean algebras.

In recent years there have been several books published, some quite good, with titles like "Discrete Mathematics for Computer Science" (e.g., Stanat and McAlister [**25**], Tremblay and Manohar [**27**], and Levy [**14**]), which focus, in the main, on topics like those above and which are aimed at second-year computer science students.

Precisely how the two-year curriculum discussed would support the first two years of computer science is beyond my scope here but is discussed in [**22**].

What mathematics should follow the first two years for computer science students? A third year of undergraduate mathematics is almost a necessity for such students; the almost obvious choices for this third year are:

1. *Calculus.* We argued previously only that calculus is hardly needed in the first two years of a computer science curriculum. But we have not argued, and would not argue, that calculus plays no role in computer science more generally. Indeed, there are numerous areas of computer science where calculus plays an important role, of which we mention only a few for illustration:

(a) One can go to considerable depth in the *analysis of algorithms* without a need for the tools of calculus, but these tools are eventually invaluable, as a perusal of Knuth's books [**12**] will exemplify.

(b) The subject matter of continuous *probability and statistics* are necessary to the practicing computer scientist in a variety of contexts (e.g., in the analysis of hardware and software performance and in the design and analysis of simulation models and experiments).

(c) Most computer scientists still view *numerical analysis* as a branch of their subject; of course the foundations of much of numerical analysis as well as its tools are dependent upon calculus.

(d) Some topics, such as *queueing theory*, that can be successfully introduced without calculus can be considered in any depth only with the help of calculus.

Finally, we note that the tools of discrete and continuous mathematics have always been interrelated in much of mathematics education. To propose the study of one to the exclusion of the other, even if one is much more relevant to a discipline than the other, would be pedantic.

Thus we argue for a semester of calculus at the beginning of the third year for computer science students. Why only a semester? For two reasons:

(a) Because a great deal more can be covered in one semester at the junior level than at the freshman level due to the presumed mathematical maturity of students who have had a two-year curriculum as outlined above.

(b) Because some topics (e.g., limits, summation of series) covered in the normal calculus sequence and related notation will have been introduced in the first two-year sequence, less time need be spent on them in calculus itself.

Thus, it seems plausible to me that at the end of a one-semester calculus course in the junior year, the student would have covered as much material as at the end of a one-year freshman sequence.

As a final word on this subject and to note that the idea of preceding calculus with discrete mathematics is not new, here is a quote from a section in [17] entitled "Thoughts on a 'Postponed' Calculus Course with Emphasis on Numerical Methods":

> In recent years many questions have been raised about the special role played by the basic calculus sequence as the first set of courses in traditional college mathematics curricula. There are many arguments for beginning with the calculus, *but with the growth of computer science and the need for more mathematics in the behavioral and social sciences there are more and more arguments for postponing the calculus courses.* [Italics added.]

2. *Statistics.* We argued above the importance of statistical reasoning to computer scientists as indeed it is important to all scientists. The second half of the junior year, following the calculus course, is the time for such a course. Except that the student would already be familiar with some discrete probability and its applications from the first year course, this one-semester course could be the one often taught by mathematics or statistics departments at the junior level.

4. Undergraduate Mathematics for Mathematics Students. Now—and most important—we consider the implications of the foregoing for undergraduate education for mathematics majors. Specifically we ask: Can—and should—the two-year discrete mathematics curriculum outlined in Section 3 be considered as an alternative to, or even as a replacement for, the traditional two-year calculus (–linear algebra) sequence?

Two motivations for considering this question suggest themselves:

1. The year of discrete mathematics outlined in Section 3 for computer science students is also a sensible one—more sensible than a year of calculus—for social, behavioral, and management science undergraduates [22]. Therefore, if the two-year sequence proposed for computer science students is also appropriate for mathematics majors, this would ease the teaching problem considerably at smaller colleges where teaching resources are limited. (Such colleges seldom have engineering programs which need service from the mathematics department. If discrete mathematics becomes the standard for the first two years of mathematics, that leaves the problem of how to handle physical science students for whom —for physics majors, at least—calculus is still a necessity in the freshman year. There is, I think, no simple short-term solution to this problem other than to teach a service calculus course for such students.)
2. More compelling, however, is the possibility that introducing potential mathematics majors to college mathematics via discrete mathematics is to be preferred on its intrinsic merits to the calculus sequence.

What reasoning can be adduced to support the second motivation?

1. We note first that teaching discrete mathematics instead of calculus to potential mathematics majors in the first year is not such a drastic change as it may at first appear. Given that mathematics majors would take a two-year discrete mathematics sequence like that outlined in Section 3, the obvious junior-level mathematics for such students would be a year of calculus (of course!). In such a year, it should be possible to cover essentially as much as in three semesters of freshman and sophomore mathematics. Thus, at the end of the junior year, mathematics majors could be as well prepared for standard senior mathematics electives as they are now. Indeed, we should emphasize that a change to discrete mathematics as the normal freshman course would, in fact, not result in students receiving a bachelor's degree in mathematics being familiar with a very different set of topics in mathematics than they are now. What it would result in is students with a quite different orientation toward mathematics (and, perhaps, computer science also) than is customary now. The implications of this for graduate study in mathematics, and for the kinds of jobs bachelor's degree recipients might obtain, are considerable.
2. At least insofar as mathematics majors themselves are concerned, the foregoing is more negative about why no harm would be done than positive about why starting college mathematics with discrete mathematics might be directly advantageous. More specifically, we ask: Does the current thrust of mathematics itself support the changes proposed? This question needs consideration from a variety of points of view:
   (a) First, we may ask whether the current thrust of mathematics research supports a greater emphasis on discrete mathematics. This is not a question which can be answered definitively given the vast and variegated fabric of mathematics research today. What can be said—and probably all that can be said—is that the level of research on discrete mathematics, in particular, and in those areas of mathematics strongly affected by computers more generally (e.g., numerical analysis, number theory) has grown much faster than in more classical areas of analysis over the past two decades [**22**]. Therefore, it is, perhaps, fair to conclude that the proposals in this paper are not in obvious conflict with the current fabric of mathematics research.
   (b) Second, we consider the professional needs of bachelor's degree recipients in mathematics who seek jobs in business or industry (but not teaching—see (c) below). Until the 1950's such jobs as there were in business and industry for those with bachelor's degrees in mathematics were mainly in scientific industry (e.g., research laboratories) or in positions where the most prevalent task was statistical data analysis. In both areas the tools of classical analysis were of paramount importance.

   But the computer revolution has profoundly changed this. A considerable number of the kinds of jobs alluded to above still exist. However, added to these have been a large number of computer-related jobs where the preponderant mathematics tools are those of numerical analysis, a mixture of classical and discrete analysis. More recently, as the effective design and analysis of computer algorithms has been seen to be increasingly important in making effective use of computers, jobs with an essentially discrete mathematics flavor have become increasingly common.

   I know of no data which separate the industrial positions open to mathematics majors by the type of mathematics most needed in them. But I think it is safe to say that the kind of program proposed herein, which is a mixture of the discrete and continuous, would prepare the mathematics major for a wider variety of jobs than the classical curriculum in which discrete mathematics typically plays such a minor role. (Of course this problem could be remedied by increasing the discrete component in the current curriculum without materially changing the first two years of that curriculum. My contention in this context is only that, in terms of job preparation, such a curriculum is not to be preferred to the one proposed herein.)

   The above ignores the fairly large number of mathematics majors who are still hired into essentially programming jobs with a rather small mathematics component. I

believe the discrete mathematics curriculum discussed here is clearly preferable to the classical curriculum for such people. But this argument should not be pressed too far because the phenomenon of mathematics students being hired into programming jobs is a transient one pending the time when computer science programs produce enough graduates to satisfy the demand for them.

(c) Third, and last, we consider the needs of those mathematics majors who embark on careers in teaching at the grammar and high school level. The important issue here, I think, is to prepare prospective teachers as well as possible for a teaching career spread over many years during which there may be considerable need for adaptation to new course material or approaches to teaching. Lack of adaptability has perhaps been one of the reasons that new approaches to teaching school mathematics have been less successful than was hoped. At least it seems to be the case that the general failure and weakness of "computer math" courses in the high schools has been due not only to a lack of familiarity with computers but also to a lack of knowledge of the mathematics—that is, discrete mathematics—most relevant to such a course at the high school level. Anyhow it seems fair to conclude that an undergraduate mathematics program better balanced between discrete and classical analysis than is now common would certainly be advantageous to prospective teachers of mathematics.

And we note that the growing importance of discrete mathematics is no transient phenomenon. It has been fueled by the growth of computers and computer science, a growth that will surely not end until computers are ubiquitous in all branches of science and technology and, more generally, in society broadly. One certain result of this is that computer science will become the largest, perhaps by far the largest, source of problems for mathematics and mathematicians.

It is also incumbent upon me to consider the argument against preceding calculus with discrete mathematics. At the heart of this argument is the belief that calculus is the natural transition course from high school to college mathematics. This is a difficult argument to deal with because of its totally qualitative nature. It is true that the experiments with "finite mathematics" for freshmen in the sixties, following the publication of the book by Kemeny et al. [**11**], seem to have pretty well died out. But this is more likely to have been because these courses were not part of an integrated curriculum than because it was inherently unreasonable to teach such mathematics to freshmen. (It should also be noted that these finite mathematics courses were considerably different from what we have proposed in Section 3.) Too often, unfortunately, such courses came to be viewed as an easier way for non-mathematics or non-science majors to satisfy a mathematics requirement than by taking calculus.

My view on this argument is expressed in part by Hammer [**9**], who advocates that "the finite calculus should be presented before the calculus" on the grounds that the latter is more easily assimilated by the student who understands the former. This point of view is similar to that of Gordon [**8**].

Even so, however, there is the question of whether discrete mathematics serves to reinforce high school mathematics as well as calculus does. This is a matter of importance, because we have all experienced the phenomenon that "students always have to be taught what they should have learned in the preceding course" [**5**]. In terms of manipulative skills, particularly those of high school algebra, discrete mathematics requires more consistent use of such skills than calculus does. As to trigonometry I see little to choose between the two. Both require some recourse to trigonometric concepts and identities. Calculus courses do, of course, spend considerable amounts of time on differentiation and integration of trigonometric functions, more time than discrete mathematics courses would spend on summation or differencing of trigonometric functions, but little of this material involves much of *trigonometric* significance. On balance, I believe it is hard to make a strong case favoring either calculus or discrete mathematics as a reinforcer of high school mathematics.

Moreover, if discrete mathematics should come to be accepted as a (or *the*) standard freshman mathematics course, this would undoubtedly suggest some changes in the high school curriculum.

Here I would note only that this "top-down" approach to changes in the secondary and primary school curricula has much to recommend it in contrast to the "bottom-up" approach which has been prevalent in recent years.

Still, when all is said and done, the pro and con arguments rest on very slender foundations with little hard evidence to support either side. At the least, it seems clear to me that the *a priori* arguments against the teaching of discrete mathematics to freshmen can hardly be considered compelling enough to suggest that experiments in this area are not warranted.

5. Halfway Houses. The foregoing has purposely contrasted two models of undergraduate mathematics: (1) the prevalent, current one in which the first two years are concerned (mainly) with the calculus, and (2) a proposed model in which the first two years would be devoted to discrete mathematics. By so doing I have been able to contrast clearly the two approaches without muddying the waters with possible compromises. But not only are such compromises possible, they need consideration, which we briefly give them in this section.

Earlier sections of this paper have stressed the arguments in favor of discrete mathematics as the subject matter for the first two years of college mathematics. By considering some of the arguments against this approach, we shall be able to define most easily some possible halfway houses between the purely discrete and the purely continuous approaches.

1. Perhaps the most serious argument against the discrete approach has been alluded to earlier. This is that the physical sciences and engineering need calculus early to support their courses and curricula. With the discrete approach, therefore, you would limit the opportunity of students to postpone the choice of a major until the end of the sophomore year, because a student who embarked on the discrete sequence in the freshman year would find it difficult to major in a physical science or engineering. (The other side of this, of course, is that students who start with the calculus sequence may, in the future, find it hard to major in computer science or even management.)
2. Even for students not majoring in the physical sciences or engineering (e.g., computer science students), a lack of calculus in the first two years may cut them off from desirable courses. Consider, for example, the student who wishes to take a course in integrated circuit design, which must be preceded by an electronics and linear circuits course, which in turn needs physics, which needs calculus. (A response to this argument is that it's true but that it must be accepted that no undergraduate program which is educational as well as professional can possibly allow the student all options in the major.)

As the parenthetical comments indicate, I think there are rebuttals to the arguments above and probably also to other arguments that could be made against the discrete mathematics idea. But certainly there is also enough merit in these arguments to make some people who are sympathetic to my general thesis nevertheless look for less radical ways to implement it. Among the various possibilities, the most obvious and straightforward would be to divide the first two years between discrete mathematics and calculus. For example:

(a) Prospective computer science majors could take a year of discrete mathematics followed by a year of calculus. For some further thoughts on this subject, see **[17]**.

(b) Prospective physical science and engineering majors could begin with calculus the first year and then take a year of discrete mathematics. In this context it should be noted that (1) the usual two years of calculus–linear algebra for students pursuing these areas is not required because *all* (or even perhaps a major fraction) of the material covered is of immediate professional need but, rather, in part at least, to develop mathematical sophistication; in this latter context one year of calculus and one year of discrete mathematics should serve just as well; and (2) because physical scientists and engineers should become sophisticated users of computers, the discrete mathematics would be of substantial use to them and would support a computer science sequence in the sophomore year.

(c) For mathematics majors, either of the above or a program of one semester each of calculus and discrete mathematics in each of the first two years could be considered. (It is interesting to note, as Professor A. W. Tucker has informed me, that a quarter-century ago the Committee on the Undergraduate Program (the forerunner of CUPM) proposed a freshman year of "universal mathematics" consisting of one semester of discrete mathematics and one semester of polynomial calculus and analytic geometry.)

(d) For those students unsure of their eventual major, one semester each of calculus and discrete mathematics during the first year could be satisfactory and would avoid the foreclosing of options.

Although, ideally, courses in both calculus and discrete mathematics would be taught differently depending upon whether or not students were having their first exposure to college mathematics; at a small college with limited teaching resources not too much need be lost if freshmen and sophomores were mixed in the same course.

Readers of this paper will surely imagine other possibilities than those described above.

6. Conclusion. Questions of education arouse strong feelings. For this reason, and because seldom, if ever, can propositions about education be proved or even strongly supported with evidence, they provoke strong statements. We have tried to avoid unequivocal statements unsupported by evidence in this paper and we hope our critics will also. Our essential proposition is simple and not immodest: *It is time to consider (i.e., try) an alternative to the standard undergraduate mathematics curriculum which would give discrete analysis an equivalent role to that now played by calculus in the first two years of the undergraduate curriculum.*

Whatever opinion anyone may have about the essential merit of this idea, it would be difficult to argue that it could do any harm to students on which it was tried. True, there are always initial difficulties with new courses and curricula. Texts are untried or unavailable. Experience is lacking on what is easy, what is hard, what seems to work, what doesn't, etc. But counterbalancing this is the enthusiasm typically present in people trying something in which they believe or, at least, for which they have high hopes, an enthusiasm which inevitably communicates itself to students and stimulates and involves them. Whether or not you judge the proposals in this paper to be a radical departure from the present curriculum, it seems unlikely that anyone, on balance, could believe that they would do educational or professional harm to the students.

Acknowledgement. I have had comments, criticism, and help from many mathematicians and computer scientists in the preparation of the paper [**22**] from which this paper has been drawn. Of these I am particularly grateful to Steven Garland, Richard Hamming, Dennis Martin, David Moursund, Gerald Myerson, Elliot Organick, Gerald Rising, Jacob Schwartz, Alan Tucker, and Albert Tucker.

References

1. ACM Curriculum Committee on Computer Science, An undergraduate program in computer science—preliminary recommendations, Comm. ACM, 8 (1965) 543–548.

2. ______, Curriculum 68—Recommendations for academic programs in computer science, Comm. ACM, 11, (1968) 151–197.

3. ______, Curriculum 78—Recommendations for the undergraduate program in computer science, Comm. ACM, 22 (1979) 147–166.

4. A. T. Berztiss, The why and how of discrete structures, SIGCSE Bulletin, 8 (1976) 22–25.

5. R. P. Boas, Jr., "If this be treason ...," this MONTHLY, 64 (1957) 247–249.

6. E. W. Dijkstra, Programming as a discipline of mathematical nature, this MONTHLY, 81 (1974) 608–612.

7. W. Givens, Implications of the digital computer for education in the mathematical sciences, Comm. ACM, 9 (1966) 664–666.

8. S. P. Gordon, A discrete approach to computer-oriented calculus, this MONTHLY, 86 (1979) 386–391.

9. P. C. Hammer, Computer science and mathematics, in Proceedings of the IFIP World Conference on Computer Education, vol. 1, IFIP, Amsterdam, 1970, pp. 53–55.

10. IEEE Computer Society Education Committee, A Curriculum in Computer Science and Engineering, IEEE Computer Society, 1977.

11. J. Kemeny, J. Snell, and G. Thompson, Introduction to Finite Mathematics, 1st ed., Prentice-Hall, 1956; 3rd ed., 1974.

12. D. E. Knuth, The Art of Computer Programming, vols. 1-3, Addison-Wesley, 1968, 1969, 1973.
13. ______, Computer science and its relation to mathematics, this MONTHLY, 81 (1974) 323–343.
14. L. S. Levy, Discrete Structures of Computer Science, John Wiley, 1980.
15. MAA Committee on the Undergraduate Program in Mathematics, A General Curriculum in Mathematics for Colleges, CUPM, 1965.
16. ______, Recommendations for an Undergraduate Program in Computational Mathematics, CUPM, 1971.
17. ______, Recommendations on Undergraduate Mathematics Programs in Computing, CUPM, 1972.
18. MAA, PRIME 80, Proceedings of a Conference on Prospects in Mathematics Education in the 1980s, 1978.
19. J. E. McKenna, Computers and experimentation in mathematics, this MONTHLY, 79 (1972) 294–295.
20. National Research Council, The Role of Applications in the Undergraduate Mathematics Curriculum, 1979.
21. H. O. Pollak, quoted in The Chronicle of Higher Education, 15 May 1978, p. 3.
22. A. Ralston, Computer Science, Mathematics and the Undergraduate Curricula in Both, Technical Report 161, Department of Computer Science, SUNY at Buffalo, 1980.
23. A. Ralston and M. Shaw, Curriculum 78—Is computer science really that unmathematical?, Comm. ACM, 23 (1980) 67–70.
24. J. B. Rosser, Mathematics courses in 1984, this MONTHLY, 79 (1972) 635–648.
25. D. F. Stanat and D. F. McAlister, Discrete Mathematics in Computer Science, Prentice-Hall, 1977.
26. O. E. Taulbee and S. D. Conte, Production and employment of Ph.D.'s in computer science, Comm. ACM, 20 (1977) 370–372.
27. J. P. Tremblay and R. Manohar, Discrete Mathematical Structures with Applications to Computer Science, McGraw-Hill, 1975.

MACHIAVELLI AND THE GALE-SHAPLEY ALGORITHM

L. E. DUBINS AND D. A. FREEDMAN
Department of Statistics, University of California, Berkeley, CA 94720

Summary. Gale and Shapley have an algorithm for assigning students to universities which gives each student the best university available in a stable system of assignments. The object here is to prove that students cannot improve their fate by lying about their preferences. Indeed, no coalition of students can simultaneously improve the lot of all its members if those outside the coalition state their true preferences.

1. Introduction. The object of this paper is to generalize the following result of Gale and Shapley [**1**]. For simplicity, suppose first that there are equal numbers of students, denoted

---

This paper describes joint research, but Freedman is mainly responsible for the exposition, including this footnote.

Before he completed his studies for his doctorate in mathematics at the University of Chicago in 1955, Lester Dubins had solved the problem of minimizing the length of a planar curve subject to boundary conditions and a curvature constraint and had settled a conjecture concerning an infinite game with incomplete information. His thesis, written under the guidance of Irving Segal, was of a more abstract nature and concerned a Radon-Nikodým derivative for Banach-space-valued random variables. After spending two years at the Carnegie Institute of Technology, he enjoyed three NSF postdoctoral fellowship years, two of which were spent at the Institute for Advanced Study at Princeton. Since 1960, he has been on the faculty of the University of California, Berkeley. His research, influenced by Leonard J. Savage and Bruno de Finetti, is largely concerned with the mathematics of (subjective and finitely additive) probability with excursions into other areas, particularly geometry, is oriented toward problems which are intuitive and concrete, and is frequently collaborative. He coauthored a book with L. J. Savage which was published by McGraw-Hill in 1965 under the title *How to Gamble if You Must* and reprinted by Dover in 1976 as *Inequalities for Stochastic Processes*.

David Freedman got his Ph.D. at Princeton in 1960; his thesis supervisor was William Feller. He spent a postdoctoral year at Imperial College, London. In 1961, he joined the faculty of the University of California, Berkeley. He has published many articles, some monographs, and an elementary textbook. His theoretical research is mainly in probability and statistics, on representation theorems like de Finetti's, on the sampling behavior of histograms, and on bootstrap methods for confidence intervals or significance levels. His applied work is in the legal area and in the validation of energy data and energy models.—*Editors*

generally by S's, and universities, denoted by U's. Suppose that each university is to admit exactly one student. (More realistic assumptions are made in Section 4 below.) Each student rank orders all the universities, and each university rank orders all students. The object is to pair the students and universities off in a *stable* way. By definition, an instability is created by two pairs, S-U and S'-U', where S prefers U' to U, and reciprocally U' prefers S to S'. Nothing is assumed about the preferences of S' and U. If there are no instabilities, the system is said to be stable. Gale and Shapley prove the existence of a stable system of assignments.

$$\begin{tikzcd} S \arrow[r, dash, dashed] \arrow[dr, dash] & U \\ S' \arrow[r, dash, dashed] & U' \end{tikzcd}$$

Each student S has an "available set" $A(S)$ of universities: the ones S can get under some stable assignment. These available sets are nonempty. Consider assigning to S that university in $A(S)$ that S likes best. Gale and Shapley prove that this assignment is one-to-one, and stable.

Here is a sketch of a proof of the Gale-Shapley results which differs from theirs in detail only, but introduces some ideas needed later. Imagine the universities (much reduced in size) lined up in a room, with the students waiting outside in a hall. One student, S, walks into the room and applies to the university S likes best: this completes move #1. Then another student walks in and does likewise; in case both apply to the same university, it keeps the preferred applicant and rejects the other, who goes back outside to the hall: this completes move #2. And so on: student S_j applying to the university S_j likes best—among those that have not previously rejected S_j.

There are two rules to observe.

(1) If there are still students outside in the hall, one, say S_j, goes into the room and applies to that university which S_j likes best, among those which have not previously rejected S_j. This initiates a move.

(2) A university with two applicants keeps the preferred one and rejects the other, who goes back outside to the hall. This completes a move.

Any sequence of moves made in obedience to rules (1) and (2) will be called a "Gale-Shapley algorithm."

(3) THEOREM. *Any Gale-Shapley algorithm terminates. At termination, the students and universities are paired off, one-to-one. This pairing is stable. And, in fact, each student S will be paired with the university S likes best in $A(S)$.*

Theorem (3) will be argued in a moment, but first a statement of the new results. Suppose a student, called Machiavelli, lies, that is, does not apply to the universities in the order of true preference. Can this help Machiavelli? The answer is no, not if the others continue to tell the truth. Similarly for coalitions of student liars. For universities, however, it is another story. These issues will be discussed in Sections 2, 3, and 4 below.

Proof of Theorem (3). Suppose there are n students and n universities. By rule (1) each student applies at most once to each university. Consequently:

(4) A Gale-Shapley algorithm terminates in n^2 moves or less.

Clearly, rules (1) and (2) imply:

(5) Each student applies to successively less desirable universities. For each university, however, the applicants look better and better.

At the end of every move, there are some students in the hall, and an equal number of universities in the room, who have not yet had applications. The remaining students and

universities are paired off, one-to-one. After a university gets its first application, it always has one. Furthermore:

(6) The algorithm ends when each university has had at least one application.

Next, it will be argued by induction that:

(7) At the end of every move, the pairing in the room is stable.

Plainly, this is so before move 1. Suppose it is so before move k, and consider the assignment at the end of that move. Now there cannot be two pairs S-U and S'-U', where S prefers U' to U, while U' prefers S to S'. For if S prefers U' to U, then S has already applied to U' and been rejected, by rule (1). Now U' must prefer the current applicant S' to the previous one S, by the fact (5). This completes the proof of (7).

$$\begin{array}{l} S\text{---}U \\ S'\text{---}U' \end{array}$$

The next point, though similar, is a bit trickier.

(8) If a student S is rejected by a university, that university is not in S's available set.

This is vacuous at move 1. Suppose it were so for moves 1 through $k-1$, and S is paired with U at the end of move $k-1$. On move k, suppose S' applies to U. Now U must retain one of these two applicants, say S_1; call the rejected applicant S_2. By way of contradiction, suppose there were a stable assignment in which S_2 got U. Now S_1 has to get some university, call it U'. At the risk of the obvious, S_1 and S_2 are different students; U and U' are different universities.

Case 1: S_1 applied to U' before move k. Then S_1 must have been rejected by U', because S_1 is applying to U on move k. So this system is unstable, by the inductive assumption.

Case 2: S_1 did not apply to U' before move k. Now S_1 prefers U to U', by rule (1). And U prefers S_1 to S_2, the proof being that it rejected S_2. Again an instability.

$$\begin{array}{l} S_1\text{---}U' \\ S_2\text{---}U \end{array}$$

To sum up, the algorithm terminates by (4); the resulting system is stable by (7); and it is optimal for the students by (8). This completes the proof of the theorem. □

2. Enter Machiavelli. One of the students—named M for Machiavelli—will now be treated differently from the rest. M has some true rank ordering on the universities, and if M participates in a Gale-Shapley algorithm following rule (1), M will get some university: the best in M's available set. This is *fair play*. But now permit Machiavelli to lie, that is, to use some false rank ordering. This is *foul play*.

(9) THEOREM. *Suppose M participates in a Gale-Shapley algorithm, but uses a false rank ordering. The university M gets by this foul play is no better—measured by M's true rank ordering—than the one M would have got by fair play.*

For the proof, imagine that M waits outside in the hall until all the others have paired off. This will be called the *prologue*. At the end of the prologue, there will be one university, call it W, which has not yet received an application. M now enters and starts applying in accordance with the rules—but using the false rank ordering. Clearly,

(10) The algorithm terminates when W gets its first application.

No generality is lost by assuming that M does not move until the others are paired off: as Theorem (3) shows, all Gale-Shapley algorithms lead to the same system of assignments. (The algorithm is now being applied with M's false rank ordering in place of M's true one.)

The main step in the proof of Theorem (9) is Lemma (11) below, which requires two definitions. A *scenario* is a sequence of applications for M—an initial segment of a rank ordering. One scenario, for instance, is specified by naming three universities:

$$A \quad B \quad C.$$

The interpretation: M applies first to A; if rejected, M tries B next; if rejected there too, M goes on to C. In general, a scenario is specified by a list of universities; no university appears twice on the list, but the list need not be exhaustive. The action called for by a scenario stops when

- either M is rejected by the last university specified in the scenario (C, in the example);

or

- the whole algorithm stops, W getting its application.

Corresponding to each scenario, there is a script that tells exactly what happens as the action unfolds, after the prologue. A script can be written in standard form as in Table 1.

TABLE 1. Standard script

Question marks indicate that the objects are undefined.

Line	University	rejects	Student	who applies to	University
0	???		$S_0 = M$		$U_0 = A$
1	U_0		S_1		U_1
2	U_1		S_2		U_2
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$k-1$	U_{k-2}		S_{k-1}		U_{k-1}
k	U_{k-1}		S_k		U_k
$\vdots$	$\vdots$		$\vdots$		$\vdots$

The table is interpreted as follows. To fix ideas, suppose again that the scenario is $A\ B\ C$.

Line 0. M enters and applies to A, and so S_0 is M and U_0 is A. Suppose A isn't W.

Line 1. U_0 now has two applicants and must reject one, say S_1. Then S_1 applies to another university; call it U_1. Of course, if S_1 is M, then U_1 must be B, according to the scenario. If S_1 isn't M, then U_1 is determined by S_1's rank order, in accordance with the rules.

Lines 2, 3 ... are interpreted in a similar way. The last line is special, and there are two cases.

Case 1: M is rejected by the last university in the scenario. Then the last line is:

Line	University	rejects	Student	who applies to	University
k	U_{k-1}		M		???

In our example, the scenario was $A\ B\ C$, so U_{k-1} is C.

Case 2: The last university W gets its application. The last line is

Line	University	rejects	Student	who applies to	University
k	W		???		???

In any case the table has finite length, by (4).

Note: For all $k \geqslant 1$, the first university mentioned in line k is the same as the last university mentioned in line $k-1$, namely U_{k-1}. In general, the same student will be mentioned several times in the sequence $S_0, S_1, \ldots$; likewise, U_i and U_j can easily be the same, even if $i \neq j$.

One more definition. Consider two scenarios, #1 and #2. Then scenario #1 is *smaller* than #2 if every university mentioned in #1 is also mentioned in #2: order is immaterial. Thus, $A\ B\ C$ is smaller than $E\ B\ D\ C\ A\ F$.

(11) THE SCENARIO LEMMA. *Suppose scenario #1 is smaller than scenario #2, and that*

(12) *M makes every application indicated in the larger scenario.*

Then every rejection and application in the script for the smaller scenario occurs, sooner or later, in the script for the larger scenario.

Proof. The argument is by induction on the line number in the script for the smaller scenario. In line 0, M comes in and applies to U_0; by assumption (12), this application occurs in the script for the larger scenario. Now make the inductive assumption:

(13) All the rejections and applications in lines 0 through $k-1$ of the script for the smaller scenario occur, sooner or later, in the script for the larger one.

Consider line $k \geqslant 1$ of the script for the smaller scenario. To avoid trivialities, suppose this isn't the last line of the table. It will be shown that the rejection and application in turn occur in the second script as well:

$$\text{line } k\ U_{k-1} \text{ rejects } S_k \text{ who applies to } U_k.$$

Line k of the script for the smaller scenario begins with university U_{k-1} rejecting student S_k. So S_k must already have applied to U_{k-1}: either in the prologue, or in lines 0 through $k-1$ of the script. If not in the prologue, this application must occur somewhere in the script for the larger scenario, by inductive assumption (13). Furthermore, according to rule (2), university U_{k-1} must have been applied to by a student preferred to S_k, either in the prologue or in lines 0 through $k-1$ of the script for the smaller scenario. If not in the prologue, this application too must occur somewhere in the script for the larger scenario. The upshot is that under the larger scenario, poor S_k must again be rejected by U_{k-1}. This event does not occur in the prologue, by assumption: so it must occur in the script.

Line k of the script for the smaller scenario ends by having S_k apply to U_k. There are two cases.

Case 1: S_k is M. This application gets made in the script for the larger scenario, by assumption (12).

Case 2: S_k isn't M. Now U_k in the script for the smaller scenario is identifiable. By rule (1), this is the university ranking after U_{k-1} on S_k's list. As shown above, S_k gets rejected by U_{k-1} in the script for the larger scenario, and must then apply to U_k.

This completes the induction, except for the last line of the table. The argument there is similar, and is omitted. □

Proof of Theorem (9). Suppose that M would get M's ith choice under fair play, where $i \geqslant 2$. By way of contradiction, suppose there is some scenario

(14) $A\ B\ C \ldots U$

that gets M a university U that M ranks ahead of i. Then the corresponding foul play script must terminate with an application to W, while M is paired with U. In particular,

(15) M makes all the applications called for in scenario (14).

There are two cases to consider.

Case 1: M truly prefers all the other universities in scenario (14) to U. To get the contradiction, the foul play scenario (14) will be compared to a fair play scenario in which M applies to M's 1st, 2nd, ..., $(i-1)$th choices in turn. By the assumption defining Case 1, the foul play scenario is smaller than the fair play one, since U ranks ahead of i. And M makes every application called for in this fair play scenario: indeed,

(16) The fair play script ends with M rejected by M's $(i-1)$th choice.

The reason is that, under fair play, M gets M's ith choice.

The Scenario Lemma (11) applies, and shows that every application in the script for foul play, including the one to W, gets made in the script for fair play. In particular, the fair play script has to end with an application to W. This contradicts (16), and disposes of Case 1.

Case 2: M truly prefers U to at least one of the other universities in scenario (14). Delete all such universities, creating a second and smaller foul play scenario. The corresponding script, by Case 1, must end with M ignominiously rejected by U. This rejection must occur in the script for the original foul play scenario (14), by the Scenario Lemma (11): condition (12) is satisfied by (15). This contradiction disposes of Case 2.

REMARK. Two scenarios that are permutations of one another are equivalent, as long as M makes all the applications in both cases.

3. Coalitions. So far, M has acted independently. What happens if M colludes with other students?

(17) THEOREM. *Suppose several students collude in a Gale-Shapley algorithm, each using a false rank ordering. They cannot all get better universities. "Better" is relative to each student's true rank ordering, and indicates strict inequality.*

The proof is an adaptation of the one for (9). Now a scenario indicates separately for each liar the sequence of universities applied to. Imagine the liars to wait outside in the hall until the honest students are all paired off with universities: this defines the prologue. At the end of the prologue, some universities have not yet had applications: their number is equal to the number of liars. Now the liars take turns in any way among themselves applying to the universities, but following the scenario. Each scenario therefore can be expanded into many scripts. To avoid complications, a student who is rejected gets to make the next application, by convention.

The action initiated by a scenario terminates when

- any liar L has been rejected by the last university on L's list

or

- the whole algorithm stops.

If the action ends according to the first possibility, no honest students can be left outside.

Note too that with several liars, and therefore several universities that have not had applications in the prologue, some applications made before the end of the script do not cause rejections. Suppose one such occurs at line k of the script. Since an honest student will be found in the hall only after a rejection, and gets the next turn, line $k+1$ of the script must have an application from a liar.

Line	University rejects	Student who applies to	University
k	U_{k-1}	S_k	U_k
$k+1$	???	S_{k+1}	U_{k+1}

Thus U_k is receiving its first application: S_k may be honest or a liar. However S_{k+1} is necessarily one of the liars.

(18) THE GENERALIZED SCENARIO LEMMA. *Suppose scenario* #1 *is smaller than scenario* #2. *Expand scenario* #1 *into script* #1, *and scenario* #2 *into script* #2. *Suppose*

(19) *In script* #2, *each liar makes every application indicated in scenario* #2.

Then every rejection and application in script #1 *occurs, sooner or later, in script* #2.

Proof. Argue by induction on the line number in script #1, as in the proof of (11). □

Proof of Theorem (17). Number the liars as $L_1, L_2, \dots$. Suppose that, under fair play, L_j's i_jth choice is what L_j would get. By way of contradiction, suppose there is a foul play script for a scenario in which L_j gets U_{jk_j}, which is strictly better than the university that L_j would get under fair play: L_j ranks U_{jk_j} above i_j. Write the scenario as follows:

$$\begin{array}{ll} L_1 & U_{11}, U_{12}, \dots, U_{1k_1} \\ L_2 & U_{21}, U_{22}, \dots, U_{2k_2} \\ \vdots & \vdots \quad \vdots \qquad \vdots \end{array} \tag{20}$$

As before:

(21) All the applications indicated by (20) get made in the foul play script.

Furthermore, by test (6),

(22) The foul play script for (20) ends with all the universities getting applications.

Again, there are two cases to consider.

Case 1: Each L_j really ranks all the universities applied to in scenario (20) as U_{jk_j} or better. This scenario will be compared to a truncated fair play scenario, but some care is needed. To begin with, consider any definite script for fair play. The liars arrive at their final universities in some order or other. Suppose (by renumbering) that L_1 applies to the i_1th choice only after L_j applies to the i_jth choice for all $j \geq 2$. Now consider the truncated fair play scenario in which

$$L_1 \text{ applies to the 1st, 2nd}, \dots, (i_1 - 1)\text{th choices, in turn;}$$

and for $j > 1$,

$$L_j \text{ applies to the 1st, 2nd}, \dots, i_j\text{th choices, in turn.}$$

By the assumption defining Case 1, this truncated fair play scenario is larger than the foul play scenario (20). Furthermore, by definition, in the specific script for fair play under consideration, L_1 gets rejected by $i_1 - 1$ while L_j is paired with the i_jth choice for $j \geq 2$. In other words, all the proposals in the truncated fair play scenario above get made. Thus, condition (19) is satisfied, and the Generalized Scenario Lemma (18) applies. The conclusion is that any application generated under the script for the foul play scenario must also be generated in the script for the truncated fair play scenario. In particular, by (22) the fair play script would have to end with all the universities getting at least one application, rather than L_1 being rejected by the $(i_1 - 1)$th choice. This contradiction disposes of Case 1.

Case 2: Some L_j really ranks at least one of the universities applied to in scenario (20) below U_{jk_j}. Eliminate all such universities from the scenario, for every liar, and expand the reduced scenario into a reduced foul play script. Case 1 applies to this smaller scenario, proving that its script terminates with some liar L_j being rejected by the last university U_{jk_j}. This rejection must also occur in the original foul play script, by the Generalized Scenario Lemma (18). Condition (19) holds by (21). □

We originally thought a stronger result might hold, namely, that if one liar in the coalition does better, another liar must do worse; as stated above, (17) only implies that if one liar does better, ano‘her liar must do no better. However, David Gale showed us that the stronger result is false.

(23) EXAMPLE. With three students and three universities, two students can form a coalition and lie: one of the liars will do better, and the other will do no worse.

The students are A, B, C; the universities are U, V, W. The true rank orderings are presented in Table 2 below: W's rank orderings are irrelevant.

TABLE 2. The true rank orderings

	University preferences				Student preferences		
	1st	2nd	3rd		1st	2nd	3rd
U	B	A	C	A	U	V	W
V	A	C	B	B	V	U	W
W	?	?	?	C	V	W	U

One script for fair play is presented in Table 3 below, with a diagram for the positions of the applicants.

TABLE 3. One script for fair play

	Applicants to		
	U	V	W
A applies to U	A		
B applies to V	A	B	
C applies to V	A	B, C	
V rejects B, who applies to U	A, B	C	
U rejects A, who applies to V	B	A, C	
V rejects C, who applies to W	B	A	C
the algorithm ends			

Now suppose B and C form a lying coalition: B's lie coincides with the truth, but C orders the universities as W, V, U. As shown in Table 4 below, C will get the same university W; but B will improve from U to V. It is worth noting that the honest bystander A also does better, going from V to U. The improvement is at the expense of the universities.

TABLE 4. One script for foul play

	Applicants to		
	U	V	W
A applies to U	A		
B applies to V	A	B	
C applies to W	A	B	C

4. Variations and Comments

(24) Theorems (3), (9), and (17) apply even when the numbers of students and universities are unequal.

Suppose, for instance, there are more students than universities. There is a new kind of instability to mention: S is paired with U and S' is not admitted to any university, but U prefers S' to S. The quick fix is to introduce some additional (fictitious) universities, ranking below the

real universities in every student's estimation. A similar trick works if there are more universities than students.

Now consider the more realistic case, where universities may admit more than one student apiece. Each university U has a quota $q(U) \geqslant 1$, and may not admit more than $q(U)$ students. In previous sections, $q(U) \equiv 1$. This condition is now dropped. The total number of places is $\Sigma_U q(U)$. If this sum is bigger than the number of students, some universities have unfilled quotas. If the sum is smaller than the number of students, some students do not get assigned to universities.

Rules (1) and (2) require only small modifications to handle this new situation. Students walk in, one at a time, and apply to the university of their choice; rule (1) remains in force. However, a university does not reject any applicants until their number first exceeds its quota: then it rejects the lowest-ranking applicant. This process too will be called a "Gale-Shapley algorithm."

(25) Theorems (3), (9) and (17) hold when each university has a quota.

The trick here is to clone the universities: make $q(U)$ copies of university U, each copy having a quota of 1. Each student rank orders the clones arbitrarily: however, if for instance Harvard is preferred to Yale, then all the Harvard clones must be preferred to all the Yale clones.

What happens if the universities make offers to the students instead of waiting for applications? To be more explicit, line up the students in the room, and make the universities wait outside in the hall. One at a time, the universities walk in and make offers of admission. A university may have more than one offer outstanding; however, the number of offers may not exceed its quota of places. A student who gets two offers rejects the one from the less desirable university, which is then free to make an offer to the next-ranking student. The cloning trick used for (25) proves

(26) Theorems (3), (9), and (17) apply when universities make offers to students; this time, it is the universities that cannot improve their situation by lying.

It may be worth while to state (3) carefully in this new context.

- There is a stable system of assignments of students to universities in which no university admits more than its quota of students. However, if the number of places exceeds the number of students, some universities will have unfilled quotas; if the number of students exceeds the number of places, some students will get assigned to no university.

- For each university U, consider the set $S(U)$ of students admitted to U under some system or other of stable assignments. If card $S(U) \leqslant q(U)$, give U all the students in $S(U)$. If card $S(U) > q(U)$, give U the $q(U)$ students it likes best in $S(U)$. This is a stable system of assignments, and optimal for the universities.

- The Gale-Shapley algorithm terminates in the system of assignments just specified.

When the students do the applying, the algorithm optimizes for students, and no student or coalition of students can all beat the system by lying. When the universities make the offers, the algorithm optimizes for the universities and no university or coalition of universities can all beat the system by lying.

(27) Example. Return to the original rules, with equal numbers of students and universities, each university admitting exactly one student, and the students making the applications. The original algorithm defined by rules (1) and (2) optimized for the students, and no student could beat the system by lying. However, universities can improve their position by lying. There is a situation involving three students A, B, C and three universities U, V, W, in which under honest play U would get its 2nd choice student; but by lying, it gets the 1st choice. The true rank orderings are presented in Table 5 below; W's rank ordering are irrelevant.

TABLE 5. The true rank orderings

	University preferences				Student preferences		
	1st	2nd	3rd		1st	2nd	3rd
U)	A	B	C	A)	V	U	W
V)	B	A	C	B)	U	V	W
W)	?	?	?	C)	U	W	V

One script for fair play is given in Table 6 below, with diagrams for the position of applicants.

TABLE 6. One script for fair play

	Applicants to		
	U	V	W
A applies to V	-	A	-
B applies to U	B	A	-
C applies to U	B, C	A	-
U rejects C, who applies to W	B	A	C
the algorithm ends			

Now in foul play, U rank orders the students as $A\ C\ B$. One script for foul play is given in Table 7 below.

TABLE 7. One script for foul play by university U

	Applicants to		
	U	V	W
A applies to V	-	A	-
B applies to U	B	A	-
C applies to U	B, C	A	-
U lies and rejects B, who applies to V	C	A, B	-
V rejects A, who applies to U	A, C	B	-
U rejects C, who applies to W	A	B	C
the algorithm ends			

(28) POSTSCRIPT THEOREM. *Suppose M would get M's jth choice under fair play. Now M lies. There is no assignment, stable for the lie, under which M would get M's true ith choice, where i is better than j.*

Proof. Suppose there were such an assignment. This assignment would still be stable if M revised the lie to make i the 1st choice. Then M could get into this university by participating in a Gale-Shapley algorithm with the revised lie: for the algorithm gives M the best available university: Theorem (3) applied to the revised lie. Now there is a contradiction to Theorem (9). $\square$

Research for this paper was partially supported by National Science Foundation Grant MCS-77-01665.
We thank David Gale, Donald Knuth, and Jim Pitman for their comments.

References

1. D. Gale and L. S. Shapley, College admissions and the stability of marriage, this MONTHLY, 69 (1962) 9–15.
2. A Gibbard, Manipulation of voting schemes: A general result, Econometrica, 41 (1973) 587–601.
3. D. Knuth, Mariages stables, Les Presses de l'Université de Montréal, 1976.
4. A. E. Roth, The economics of matching: Stability and incentives, Faculty working paper no. 744, College of Commerce and Business Administration, University of Illinois, 1981.
5. M. Satterthwaite, Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions, Journal of Economic Theory, 10 (1975).

ON THE DEVELOPMENT OF LOGICS BETWEEN THE TWO WORLD WARS

I. GRATTAN-GUINNESS
Middlesex Polytechnic at Enfield, Enfield, Middlesex EN3 4SF, England

1. Introduction. Logic is a disparate topic, occurring in almost any field of human activity without appearing to have much character of its own. Traditionally it was associated largely with methods of reasoning and regarded as encapsulated in the principles of syllogistic logic. It was the concern mostly of philosophers and developed in the context of rather general questions. During the second half of the nineteenth century there was especial concern with connections with psychology. For example, some authors maintained that psychology is a descriptive theory concerned with how we think, while logic is a normative discipline about how we ought to think (see Richards *1980* for the case of Mill).

Mathematics began to play a significant role in logic with Boole's work around 1850 on Boolean algebra. From the technical point of view his work increased the scope of reasoning; and his contemporary de Morgan and successors Peirce and Schröder moved still further beyond the confines of syllogistic logic when they developed a theory of relations in their algebraic logic. Otherwise, however, the generality of concern remained; for example, Boole saw his own logic as concerned with the workings of the mind.

A different tradition was instituted around 1880 with the 'mathematical logic' of Frege and Peano, and its development by Russell and Whitehead. Not only is the 'logic' itself rather different in form; the motivations lie in specific questions in the foundations of mathematics, in contrast to the general concerns indicated above for other logicians. There were two principal motivations: the foundations of arithmetic; and the formal language required to express mathematical analysis in the style of Weierstrass, with especial concern for the set theory of Cantor. In Russell and Whitehead *both* problems are treated, because they followed the principle (though not the methods) of Cantor in seeking a foundation for arithmetic in set theory.

All these developments belong to the pre-history of my topic, and I shall not dwell on them here.[1] I must indicate now, however, that there was already in that period something of an overthrow of the old tradition by the new one. This paper, for example, is almost entirely concerned with the consequences and further advance of mathematical logic. I shall conclude this introductory section by indicating the topics to be discussed and also those which I have avoided or set aside.

Section 2 concerns the development of logicism, which had been expounded in Whitehead and Russell's *Principia mathematica* of 1910–13 (*PM*). Discussion is confined to Wittgenstein, Ramsey, and Quine, who made substantial revisions in, or use of, logicism. Section 3 charts the progress of Hilbert's formalism in the 1920's, including the blow dealt to it by Gödel's incompletability theorem of 1931. Section 4 surveys the interest taken in recursion and computability by Church, Kleene, and Turing, especially as an extension of the proof-method used by Gödel in the

[1]See, for example, Cavaillès *1937* and *1938*, my *1977* and *1980*, Dauben *1979*, and Bunn *1980*. Better-known but less reliable sources include Jörgensen *1931*, Bowne *1966*, and Styazhkin *1970*. Van Heijenoort *1967* contains English translations, with commentaries, of many major papers up to 1918.

I. Grattan-Guinness took an M.Sc. (Economics) in mathematical logic and the philosophy of science at the London School of Economics and has both a Ph.D. and a D.Sc. in the history of science from London University. He is currently Reader in Mathematics at Middlesex Polytechnic, England. He has been editor of *Annals of Science* since 1974 and is the founder-editor of the new annual *History and Philosophy of Logic*, of which the first volume appeared in 1980. He is also an associate editor of *Historia Mathematica* and belongs to the editorial boards of new editions of the writings of C. S. Peirce and B. Russell. He is a member of the Executive Committee of the International Commission on the History of Mathematics. His latest book, of which he was both the editor and a contributor, is *From the Calculus to Set Theory, 1630–1910: An Introductory History* (1980). He is currently working on the history of mathematical physics from 1790 to 1840, and of mathematical logic and the foundations of mathematics from 1870 to 1930.

proof of his 1931 theorem. Section 5 deals with the intuitionism of Brouwer, including some remarks on his polemics with Hilbert. Section 6 covers certain extensions to the traditional conception of mathematical logic that were proposed and/or rejected by Hilbert, Skolem, and Zermelo. Infinitary and second-order logics were the principal candidates for attention. Section 7 selects some of the topics studied by Polish logicians, concentrating on the comments on logicism by Chwistek, Leśniewski's logical systems, Tarski's contributions to semantics and related topics, and Łukasiewicz's advocacy of many-valued logics. Finally, in Section 8 I compare the situation around 1939 with that in 1918. The bibliography includes the names and dates, where known, of principal figures; the dates used in the text are dates of publication.

Even this range of topics far from exhausts the developments in logic between the two world wars. For reasons of both space and prudence I have omitted topics such as inductive logic, phenomenological logic, and the role played by logic in the development of quantum mechanics. I have also touched only *very* lightly on related topics in mathematics which bore also on logic in various ways; abstract algebra, model theory, and especially axiomatic set theory and transfinite arithmetic, although in some sections I comment on the views held on the relationship between logic, mathematics, and set theory.[2] Throughout I have had to elide many of the fine distinctions that in fact played some role in the developments, and also omit many of the historical details. For this I crave the indulgence of specialists.

2. Logicism and Its Critics. The philosophical position expounded in Whitehead and Russell's three-volume *Principia mathematica* of 1910–13 was that pure mathematics is derivable *solely* from logical principles and by logical processes of reasoning. By 'logic' they understood a form of mathematical logic comprising calculi of propositions and predicates (or propositional functions), including multi-order quantification over individuals and predicates, and certain rules of inference. This logic was intended to embrace set theory. By 'pure' they intended to emphasise the logical form '$p \to q$' for mathematical propositions. By 'mathematics' they seem to have intended to cover the "whole" subject, though the details were confined to finite and transfinite arithmetic, the foundations of real-variable analysis, and various other topics preliminary to a fourth volume on geometry that Whitehead never completed. To avoid the paradoxes of set theory the logical system included a theory of types. Unfortunately in order to express the required mathematics they had to introduce 'the axiom of reducibility', which asserted that to any predicate of an object a there exists a logically equivalent predicate which presupposes only objects of the type of a. The "logical" status of this axiom was, to say the least, not obvious.

I shall confine myself in this section to later considerations of the broad principles of logicism, rather than the various technical improvements that were rendered. Russell's conception of logic was all-embracing in that, to use later terminology, it included indiscriminately both logical and metalogical topics. Thus it was, strictly speaking, *impossible* for Russell to talk about his logic. Among other drawbacks, this situation threatens logicism with circularity; for one might assert that pure mathematics is derivable solely from logic, but also specify logic so as to ensure that pure mathematics is encompassed.

This problem was tackled by some of Russell's followers, who tried to define, or characterise, logic in a manner independent of logicism (see my *1979a*). For example, Wittgenstein announced in the preface to his *Tractatus* that his purpose was 'to set a limit [...] to the expression of thoughts' (*1922*, 3). In his work he outlined the idea of treating complex propositions as truth-functions of elementary propositions, the truth-functionality to be expressed by truth tables

[2] The general history of inter-war logic has not been written. Dumitriu *1978* and Guillaume *1978* provide brief notices of various aspects, while Hermes *1966* surveys the period 1890–1965 at similar speed. Mangione *1972–76* is much more substantial. Some aspects of developments in the 1930's are summarised *passim* in Mostowski *1965*. Van Heijenoort *1967* also contains English translations, with commentaries, of major papers up to 1931. Church *1952* contains historical notes *passim*, and his *1936–38* provides a detailed bibliography of symbolic logic up to that time; a more extensive bibliography up to recent times is in preparation by J. M. B. Moss for the Clarendon Press 'Oxford logic notes' series. References to secondary literature on specific topics are given *in situ* in later sections.

(*ibid.*, prop. 4.31). Logical propositions are characterized as 'tautologies' (6.1), that is, propositions whose truth-value is 'true' under all truth-values of their component propositions (4.46). Mathematics 'is a logical method' (6.2), and so presumably also has tautologous status.

In his introduction to *Tractatus*, Russell seized on Wittgenstein's remark about setting a limit to the expression of thoughts, and admitted the possibility 'that there may be another language dealing with the structure of the first language, and having itself a new structure, and that to this hierarchy of languages there may be no limit' (Wittgenstein *1922*, xxii). It is curious that when beginning to prepare the second edition of *Principia mathematica* a year later, Russell did not start from this important remark, which was made at the time when the recognition of metalogic as independent of logic was far from widely known. (Frege's distinction of object- and meta-language is the most explicit case of the period.) In his new material he used other suggestions of Wittgenstein, especially the idea of truth-functionality. It made his system much more extensional (that is, based on regarding collections as composed of their members rather than as defined by (intensional) properties), although the extent to which extensionality is taken does not seem to be clear.

This move towards extensionality was followed with enthusiasm by Ramsey, who had read the proofs of Russell's new material. In his own work, especially *1926*, Ramsey defined universal and existential quantification as infinite conjunctions and disjunctions of propositions, respectively; for example,

$$(\exists x)\cdot\phi x = \phi a \vee \phi b \vee \cdots \qquad \text{Df.}$$

He took all predicates as extensional, and in these terms re-structured the type theory of *Principia mathematica* without having to use the axiom of reducibility.

Ramsey also distinguished the paradoxes into mathematical ones, concerned with sets and numbers, and semantic ones, which deal with notions such as definability; and he divided type theory correspondingly into 'simple' and 'ramified' parts. Since his time it has become customary to regard semantic paradoxes as irrelevant to logicism.

Logicism also attracted the attention of the 'Vienna Circle' of philosophers, especially Carnap, who wrote one of the first books *1929* in German on logicism. In his *Der logische Aufbau der Welt* (*1928*), which outlined much of the philosophical programme of the Circle, he made considerable use of Russell's logical techniques, including the Russell/Wittgenstein view of extensionality. Interestingly, in his preface to the English edition of the book, he admitted that his use of extensionality 'is unclear in some points' (*1967*, ix).

Another Vienna Circle member took up *Principia mathematica*; for Gödel's incompletability theorem of 1931 showed that the Russellian logicist programme cannot be executed (see the next section). Thereafter the question of the relationship between mathematics, logic, and set theory was obviously wide open. Quine's work has been particularly influential in this area; he devised various logical systems of the scope of *Principia mathematica* without espousing logicism. Two of his early systems were the so-called 'New foundations' (*1937*), in which type theory is replaced by a stratification of formulae; and 'Mathematical logic' (*1940*), where the paradoxes were avoided by following von Neumann's idea of denying that certain objects can be members of sets. In these and other systems some status was assigned to set theory which is autonomous of logic. No definitive view of the distinction between logic and set theory has been laid down (see Quine *1969*); but Russell's logicist standpoint, that set theory is *part* of logic, was abandoned.

Russell himself wrote little on logicism after the second edition of *Principia mathematica*. Whitehead sketched a revised logico-arithmetical system in *1934*, based on the notion of instantiation of objects; despite many readings, I have not been able to see how far it can be developed. During the inter-war period logicism fell quite substantially in reputation; although it influenced logicians both as an early example of a comprehensive mathematico-logical system and as a source of techniques, the philosophical position itself won few followers.

I conclude this section with an event of our period which was unknown at the time but which is of especial historical interest. The first logicist was Frege, although he did not include "all"

mathematics in his viewpoint. His work was not widely known during the inter-war years, although Scholz hoped to publish his manuscripts. In one of these texts, *1925*, Frege abandoned logicism entirely, on the grounds that logic alone could not provide objects for which properties such as equality or set-membership can be appraised. The rejection of a life-long position is a rare achievement in mankind. It reveals a special kind of greatness.

3. Formalism and Its Fate. Hilbert strongly advocated the use of axiomatics in mathematics in his early years. These were the 1890's, when such a view was far from widespread among mathematicians (see Cavaillès *1937*). In the 1900's he applied the approach to the foundations of logic and arithmetic; but he was very unclear on the distinction between a formal system and its interpretation, so that his early papers make peculiar reading. Perhaps for these reasons he seems to have set the work aside; but he resumed his interests during the First World War (see his *1918*), and he developed his ideas much further in the 1920's to espouse 'formalism', in which a mathematical theory is axiomatised and treated as a string of symbols for the purpose of studying, in metamathematics, properties such as consistency, completeness, and the existence (or not) of decision procedures (that is, procedures to show that any well-formed formula of a formal system is provable or not).

Here is a simple example of his approach, taken from his paper *1922*. He gave an axiomatisation of arithmetic, roughly like the Peano axioms without the induction axiom, together with the *modus ponens* rule of inference. He proved as a lemma that a provable formula need not contain the conditional connective '$\rightarrow$' more than twice. He then showed that the system was consistent by showing that an equation $\alpha = \beta$ and its negation $\alpha \neq \beta$ are not provable.

During the 1920's Hilbert and others, especially his assistants Ackermann and Bernays, obtained results on consistency and completeness, and he and Ackermann published a textbook *1928* on mathematical logic outlining their principal interests. Among specific results, von Neumann showed in *1927* that a certain part of first-order arithmetic was consistent; Gödel proved in *1930* the completeness of the first-order predicate calculus. In the same year Herbrand studied various properties equivalent to the provability of a formula in such a system (see his *1930*), and stated (with a defective proof) a 'fundamental theorem' with the aid of which he could appraise (in *1930* and also *1931*) the consistency of, and decision procedures for, various logical and arithmetical systems. He also proved the deduction theorem for the first-order predicate calculus (*1930*, 108), which states that if a formula B is provable from A in the calculus, then $(A \rightarrow B)$ is also provable. This theorem has become one of the most widely used results in metamathematics.

Metamathematics can itself be studied in metametamathematics, and so an hierarchy of theories is erected. Although Hilbert confined his detailed studies to (first-level) metamathematics, his hope may have been that each tier in the hierarchy would become successively simpler in content and assumption, so that at some suitably high tier uncontroversial assumptions ($0 \neq 1$, say) would guarantee its consistency, and thereby transmit the guarantee down through the lower levels. By this means he could establish the consistency of mathematical theories, and thus show them to be cleansed of paradoxes.

Unfortunately Gödel's incompletability theorem *1931* showed that this conception of the relationship between tiers was incorrect. Gödel defined an axiomatic system P for first-order arithmetic, using a logical system like that of *Principia mathematica* and the Peano axioms for arithmetic, and proved that P was incompletable in that it contains a proposition A for which neither A nor not-A is provable. As a corollary he then showed that a proof of the consistency of P could not be expressed within P, but would require a formally richer system.

Although Gödel stated that the corollary 'do[es] not contradict Hilbert's formalist standpoint', on the grounds that there may be finitary proofs of consistency not expressible within P (*1931*, 615), such a possibility is very unlikely, and it was soon recognised that Hilbert's hopes for a consistency proof for mathematical theories along the lines described above, and belief that truth is equivalent to deducibility, must be abandoned. The incompletability theorem itself also seems

to have affected logicism, for P could be converted to a system like *Principia mathematica*, which was thus also shown to be incompletable.

While Gödel's theorem rebuffed Hilbert's hopes, it did not detract from interest in metamathematics; indeed, it led to interest in non-finitary consistency proofs. Of particular note was Gentzen's proof of the consistency of first-order arithmetic in *1936*, in which he made use of transfinite induction. Gentzen made other important contributions to metamathematics in his dissertation *1935*. He proved a 'principal theorem', which has some similarity to Herbrand's 'fundamental theorem' mentioned above and was also of value in producing consistency proofs and seeking for decision procedures. He also recast the predicate calculus in a manner which is now called 'natural deduction' and stands closer to heuristic reasoning than to traditional axiomatisations of the calculus. The contributions of Gentzen and Herbrand exercised a marked influence on the development of post-Gödelian metamathematics. Indeed, metamathematics became one of the chief interests of logicians in the inter-war period.

4. Recursion and Computability: American Logic. In order to prove his incompletability theorem Gödel had to design both his formal system P and his metamathematical concepts so that his metastatement of the incompletability of P could be expressed in its own formal language. Since P was a formulation of first-order arithmetic, the metamathematics had to be expressible in arithmetical terms. This led Gödel to his process of the 'arithmetisation of metamathematics', in which he developed a theory of what are now called 'primitive recursive functions'.[3]

Recursion had already been noted in the 1920's, especially in the metamathematical studies discussed in the previous section; Skolem *1923b* is also a notable contribution. But Gödel's treatment was the most systematic hitherto, and his theorem suggested that there may be similar limits to recursion. These researches were actively pursued in the 1930's. Many of the principal papers are reprinted in Davis *1965*, and a detailed account of the results is provided by Kleene *1952*, pt. 3.

We take as initial functions the successor function ($\phi(x) = x + 1$), the constant function ($\phi(x) = K$), and the identification function ($\phi(\{x_i\}) = x_r$). A primitive recursive function is defined as obtainable from a finite number of uses of the initial functions and of schemata given by

$$\phi(x) = \theta\big(\{\psi_i(x)\}\big) \quad \text{and} \quad \phi(0) = k, \quad \phi(y+1) = \psi(y, \phi(y)).$$

I have stated the definition for a function of one variable; generalisations to functions of several variables are obvious.

Examples of recursive functions that cannot be obtained by the process of primitive recursion had been known since the late 1920's, and extended definitions were proposed. One, due to Kleene *1936*, defined a function as 'general recursive' if it were obtained from the processes of primitive recursion and also an evaluation procedure of the form (for functions of one variable)

$$\phi(x) = \text{the smallest value of } y \text{ such that } \psi(x, y) = 0.$$

ψ is a given function, and at least one value of y is assumed to exist. The function is not reducible to the others because no upper bound is set on the value allowed for y.

Church proposed as a thesis that general recursion be taken as the definition of the effective calculability of a number-theoretic function, and proved theorems of a generalised Gödelian type on the lack of a decision procedure for first-order arithmetic (see Church *1936*). In the same year Turing proposed as a thesis that computability be defined as the operations that can be executed by a 'Turing machine', which is the conception of a computer reduced to essentials (see his *1936*).

[3]The process of arithmetisation requires the distinction between a formal system and its interpretation to be made very carefully. Professor J. Barkley Rosser, who contributed to the early studies of Gödel's theorem, has told me in reminiscence that not until this theorem was published did logicians realise *how* carefully they had to make the distinction.

It was shown that general recursion was equivalent to computability for number-theoretic functions. Thus an interesting connection between logic and computing was established, and has continued in various forms ever since. Both of these notions were shown to be equivalent to a concept called 'λ-definability', introduced by Church as part of his proposal (initiated in *1932–33*) for a logic without variables. This proposal led to the early development of combinatory logic (see Curry and Feys *1968, passim*).

All the work described in this and the last section was basically inspired by Hilbert's programme for metamathematics, and in their treatise *1934–39* he and Bernays gave a detailed account of the developments. Naturally many of the results were obtained by Hilbert and his colleagues; but Americans also played a prominent role, especially Church, Kleene, Post, Rosser, and Gödel (resident in America from 1933).

Since that time America has been a leading country for logic, and saw the founding in 1936 of the Association for Symbolic Logic. This organisation is still the only international organisation for the subject, and its *Journal of symbolic logic*, also founded in 1936, was the first journal devoted exclusively to logic and related topics. By 1939 the Association had achieved a membership of around 200, which was also around the number of pages published in each volume of the *Journal*. It was founded soon after the establishment in America of the Philosophy of Science Association in 1934, together with the journal *Philosophy of science*. The Association was then strongly influenced by Vienna Circle émigrés resident in America, and so showed some interest in logic and its applications, after the spirit of Carnap's employment of techniques from *Principia mathematica* that I noted in Section 2.

A similar recognition of institutional change may be noted in the *Jahrbuch über die Fortschritte der Mathematik*, the leading journal between the two world wars for mathematical reviews. It had established a section for set theory before 1918, containing both the abstract and the point-set branches, and also a rather vague section called 'Philosophy' where logic was usually covered. But in volume 61, for 1935 and published in 1936, point set theory was placed within 'Real functions' or 'Topology', 'Philosophy' was abandoned, and a new section entitled 'Foundations of mathematics: Abstract set theory' was established. With volume 65, for 1939 and published in 1943, this section had become 'Foundations of mathematical logic' with the sub-sections 'Philosophical', 'Logic', and 'Foundational arithmetic and algebra'; and abstract and point set theory had been reunited again as a section under 'Analysis'. Other minor changes occurred until the final volume 68 for 1942, published in 1944; but the sub-section for logic was retained. The *Zentralblatt für Mathematik* allowed for 'Foundational questions, logic' from its inception in 1931, and *Mathematical Reviews* included 'Foundations' (initially 'Foundations of analysis') when it began to appear in 1940.

5. The Emergence of Intuitionism. The intuitionistic philosophy of mathematics is associated primarily with Brouwer's rejection of the law of excluded middle as a valid logical principle. He introduced his ideas in the 1900s, and first largely applied them to Cantorian set theory; but around 1918 he began to develop them in more detail as a general viewpoint for both logic and mathematics.

Brouwer advocated that mathematical proofs be restricted to constructive processes. Thus there was a direct clash with Hilbert's formalism; for while Hilbert assumed the law of excluded middle as part of his means of proving the consistency of mathematical theories, Brouwer felt that 'the (contentual) justification of formalistic mathematics by means of the proof of its consistency contains a vicious circle, since this justification rests upon [...] the (contentual) correctness of the principle of excluded middle' (*1927b*, 491). He came to reject Cantor's theory of the actual infinite; but Hilbert supported Cantor, and used metamathematical techniques to try to prove Cantor's 'continuum hypothesis'. Brouwer also tried to replace the classical formulations of mathematical analysis with intuitionistic forms (see, for example, *1927a*).

However, Brouwer did not help his cause by the notoriously obscure formulation of his basic notions (for helpful explanations, see van Heijenoort *1967*, 446–457), and dogmatic assertions,

such as mathematics being independent of both language and logic. The discussion between him and Hilbert was largely unprofitable, partly for these reasons and partly for a clash of personalities.[4] Formalism attracted much the greater support in the inter-war period, although several logicians wrote on various aspects of intuitionistic logic and mathematics, and Weyl was an early convert to a form of intuitionism (see especially his *1927*). Heyting was the most prominent follower of intuitionism, publishing as *1934* the first textbook presenting a form of intuitionistic logic.

Although intuitionism owes its origins chiefly to Brouwer, it also suffered between the wars from his perplexing presentations and polemic style. But it has been one of the most interesting and far-reaching criticisms of the foundations of mathematics, and has continued to claim the attention of logicians to this day.

6. Beyond First-Order and Finitary Logics. The accounts in Sections 2–5 have been of specific topics and developments. In this section I wish to draw attention to an "under-current" in inter-war logic, a matter which was not quite a topic in the usual sense but an *issue* which arose and fell at various times, especially in the 1920's: the extension of logic beyond its first-order scope and finitary expression. To discuss the issue fully, I would have to go rather deeply into the development of model theory and axiomatic set theory, which would both imbalance and unduly extend this paper; I refer to Goldfarb *1979* and Moore *1981* for fuller accounts.

I mentioned in Section 2 that Ramsey urged the extensional interpretation of quantification in terms of infinite conjunctions and disjunctions. Whether wittingly or not, he was following the traditional interpretations of quantifications in algebraic logic. Schröder, in his theory of relatives, wrote quantifiers as '$\Pi_i a_{ij}$' and '$\Sigma_j a_{ij}$', for example, where a_{ij} is (in effect) a relation between variables i and j (see my *1975*) and 'Π' and 'Σ' are intended to convey the idea of infinite conjunctions and disjunctions of propositions.

During the First World War Löwenheim wrote a paper *1915* on the first-order predicate calculus following Schröderian principles. He used infinitely long expressions in logic following the Schröderian interpretation of quantification, and noted the conversion of formulae into 'normal forms' (where the quantifiers are all collected together, in certain orders, at the front of the formula). He also proved a form of 'the Löwenheim-Skolem theorem', which states essentially that every satisfiable formula of the calculus is satisfiable in a finite or denumerable domain.

Löwenheim's paper stimulated various studies in the 1920's. Skolem dealt in more detail with normal forms in a paper *1920*,[5] and produced his own proof of the Löwenheim-Skolem theorem

[4] These clashes came to a head at the International Congress of Mathematicians, held at Bologna in 1928. Brouwer, a Germanophile although born Dutch, shared the sentiments of some Germans that their country should not participate. However, Hilbert led the German delegation; and in the following year he removed Brouwer from the editorial board of *Mathematische Annalen* (see Reid *1970*, 184–188). The whole issue undoubtedly involved ideological considerations; but some aspects of the early work of Hilbert and Brouwer suggest that, concerning Brouwer *himself*, there may have been purely personal elements. Hilbert's early papers of 1904 on metamathematics are scarcely intelligible, for want of a careful distinction between symbols and their referents. Brouwer's thesis of 1907, by contrast, makes the distinction very clearly (see Brouwer *Works*, 61, 70, 101), although he does not develop a *theory* of 'mathematics of the second order'. Hilbert will probably not have read this Dutch thesis; but in later papers Brouwer reported communicating these ideas in conversation with Hilbert in 1909 (see, for example, Brouwer *Works*, 410). When Hilbert started again on metamathematics in 1917, the distinction is clearly made. Brouwer, an expert at nurturing grudges, could well have felt (with or without justice) that his ideas were being used without acknowledgement, and sought some kind of personal revenge at an opportune moment. Their personal relations had become further complicated by other personal factors (for example, Weyl's switch from formalism to a type of intuitionism).

[5] The reduction of formulae to normal forms became a very useful technique in both mathematical logic and metamathematics, as developed by Hilbert and others in the ways described in Section 3. One feature was the fact that to a given normal form there is a dual form. Here again influence from algebraic logic is evident; for duality was often used by Schröder, whereas it is largely absent from the predicate calculi of Frege and Russell (see my *1975*, 121).

using only denumerably long expressions rather than the non-denumerably infinite lengths employed by Löwenheim. Later he gave a new, but rather weaker, proof of the theorem, using only finitely long expressions, and also pointed out as a consequence what is now known as 'the Skolem paradox': that Zermelo's axiomatisation of set theory can be satisfied in a denumerable model even though the theory allows for non-denumerably infinite sets (see Skolem *1923a*).

Skolem's new proof, unlike his old one, avoided using the axiom of choice. This axiom had emerged in the 1900's in the context of Cantorian set theory and mathematical analysis, and its non-constructive character and occasional avoidability led to a lively controversy over its forms and necessity (see Sierpiński *1918*, my *1977 passim* and Moore *1980*). By the end of the First World War the controversy had largely died down, although Skolem noted the continuing reservations over the axioms—reservations which he did not share—at the end of his *1923a*. Another mathematician who had no qualms was Hilbert, who now put it to a new use; the interpretation of quantification.

In *1923* Hilbert introduced what he called 'the transfinite axiom', which asserts that if an operator τ on the predicate $A(x)$ selects the value c of x for which $A(c)$ is provable, then $A(y)$ is provable for all values of y. He advocated the use of this axiom to define universal quantification in a finitary form, in contrast to the infinitary interpretation as infinite conjunctions of propositions; existential quantification could also be defined by applying τ to not-$A(x)$. Later he restated his transfinite axiom in terms of the axiom of choice in a way which in effect interprets the existential quantification '$(\exists x)A(x)$' in terms of using a transfinite choice function to select a (or the) value of x which satisfies A (see his *1926*).

At this time Hilbert also urged the extension of logic from first-order to second-order predicate calculus (he used the terminology 'restricted' and 'extended functional calculus'), where quantification is allowed over predicates A as well as individuals x, for the purpose of expressing arithmetical notions in logical form (see, for example, Hilbert and Ackermann *1928*, ch. 4). In this respect (though not in philosophy, of course) he was moving towards the logicists' practice, for they had defined arithmetic notions in their systems and freely allowed second-order quantification. Indeed, they had admitted quantification to arbitrarily high orders; in *Principia mathematica* type theory in effect separated out the various levels of quantification.

Thus Hilbert's position was to be in favour of second-order logic but opposed to infinitary logic. Skolem, reluctant to rely on the concepts and techniques of set theory, was opposed to both these forms of logic; but in the 1930's Zermelo advocated them both. He supported second-order logic as indispensable for the formalisation of axiomatic set theory; and he used infinitary logic in a sketched theory of infinitely long proofs (the fullest account is in *1935*), from which he hoped to be able to show that all true formulae are "provable" in his extended sense of proof.

Zermelo's view seems to clash with Gödel's incompletability theorem, which was published after he began to develop his ideas. The two men spoke in September 1931 at the annual meeting of the *Deutsche Mathematiker-Vereinigung* on their very different viewpoints, and they corresponded soon afterwards (in letters published in my *1979b*). In their letters they basically re-stated their positions; in particular, Gödel continued to wish to use only finitary proofs.

Zermelo's ideas were too vague for practical development, and infinitary logics gained little interest until well after the Second World War. Second-order logic maintained some supporters, especially Hilbert, although the relationship with first-order logic was not widely discussed. The considerable differences between the two logics are often not realised even today.

7. The Rise of Polish Logic. At the end of Section 4 I noted the emergence of America in the mid 1930's as an important centre for logic. The other—indeed, the first—major national development between the wars was the rise to prominence of Polish logicians in the 1920's. The principal figures were Adjukiewicz, Chwistek, Kotarbiński, Leśniewski, Łukasiewicz, Słupecki, Sobociński, Tarski, and Wajsberg, although there are many other significant names that one could adjoin.

The rise of Poland to prominence—a position which it has retained to this day—is one of the most remarkable features of the history of twentieth-century logic; and its early development is

still more interesting because of the simultaneous emergence of a school of Polish mathematicians (see Kuratowski *1980*). Banach, Kuratowski, Lindenbaum, Mazurkiewicz, and Sierpiński were perhaps the most prominent mathematicians; and since many of them were strongly interested in algebra and set theory, their work overlapped in part with that of the logicians. Indeed, until 1928 Łukasiewicz and Leśniewski were members of the editorial board of the Polish journal *Fundamenta mathematicae*, launched in 1920 and devoted to set theory and related topics, and some logical papers appeared there. Otherwise the logicians published in various Polish academy and society journals, and also sometimes abroad.

In this section I can indicate only some principal interests; a more detailed account is provided in Jordan *1945*. A selection of English translations of Tarski's works will be found in his *1955*, and of other authors in McCall *1967*. At the time many of their works were published in French or German (or occasionally English), though some appeared only in Polish.

Naturally, in the early 1920's *Principia mathematica* was a major interest. Chwistek's contributions were rather similar to those of Ramsey, which I noted in Section 2; indeed, he anticipated Ramsey to some extent (see my *1979a*). In a two-part paper *1924–25*—rejected by *Fundamenta mathematicae* (see my *1979a*) but published by the Polish Mathematical Society—he divided Russell's type theory into 'simple' and 'branched' (that is, ramified) parts, although he was less clear than Ramsey on the distinction between kinds of paradox. He also tried to re-construct type theory while avoiding the axiom of reducibility , although his new system is harder to comprehend than is Ramsey's; for example, the extent of his commitment to extensionality is less clear. His work is of some note as an early attempt to bridge the gap between the prevailing philosophies of mathematics; for while he constructed a system of scope comparable with that of *Principia mathematica*, he also tried to specify his formal system in formalist style.

Russell also influenced Leśniewski, although here it was Russell's paradox (of the set of all sets which do not belong to themselves) which was the prime source of inspiration (see Sobociński *1949–50*). Leśniewski came eventually to construct three logical systems (see Luschei *1962* and Rickey *1977*). 'Ontology' is a modernised version of traditional logic; in its structural aspects it includes a theory of classes and relations. 'Mereology' is a study of the part-whole relationship between objects. Both these systems took as their logical basis 'protothetic', a propositional calculus assuming only equivalence as a primitive connective but also using quantification over propositions.

This 'equivalential calculus' was one of Tarski's early contributions to logic; in fact, it was the subject of his doctoral dissertation *1923*. But his main work lay in the growing interest in metalogic, especially concerning semantical questions. Here he showed some influence from Leśniewski, who in the early 1920's laid emphasis on semantic categories (see his *1929*) and on the role of definitions in formal theories and their formulation as equivalences. Tarski's most celebrated work in this area is his treatise *1933* on the semantic definition of truth as the property of a sentence in a language. He also broadened Hilbert's conception of metamathematics into a study of 'the methodology of the deductive sciences' (see *1930*), where the notion of logical consequence and Herbrand's deduction theorem (for which Tarski claimed priority) played major roles. In his text-book *1935* he laid great emphasis on the theory of deduction.

Tarski also worked with his teacher Łukasiewicz on many-valued logics. This topic received little attention between the wars (see the historical outline and bibliography in Rescher *1969*); in fact, the Poles were the logicians most active in it, with Łukasiewicz as the chief protagonist. After early work on the propositional calculus, especially as presented in *Principia mathematica*, he extended the method of truth-tables to three-valued logic, and in a joint paper *1930* with Tarski he explored many-valued logical systems. Łukasiewicz also introduced the Polish system of notation for logic, in which certain conventions are employed to avoid the use of brackets or dots in formulae.

As in America, this burgeoning interest led to organisational consequences in the mid 1930's. The periodical *Studia philosophica* was launched in 1936 to cover both logic and philosophy; it contained Tarski's German translation of his treatise *1933* on truth in its first volume. In 1939 *Collectanea logica* was planned, with a first volume on proof, as a forum for publishing Polish

logic; but the print-shop was destroyed in the bombing of Warsaw, and only offprints of some papers that had already been posted to foreign logicians and philosophers have survived.

8. Conclusions and Comparisons. In this final section I compare the situation around 1918 with that in 1939. I do so from seven points of view, the first three illustrating contrasts, the next three similarities.

8.1 *Philosophies of mathematics.* In 1918 logicism was the most fully developed philosophy of mathematics (and *Principia mathematica* the best-known logical system), and it held the dominant position. But during the next decade intuitionism and especially formalism grew to prominence. The disputes between the three schools gave the impression that they were both mutually inconsistent and also exhaustive of possible views. But during the 1930's this impression dissolved, as each school faced great difficulties or was revealed as inadequate in some ways.[6] From that time until today there was no *dominant* philosophy of mathematics, or logical system. In particular, logicism, having been centre stage in 1918, was very much in the wings from the 1930's on. Hilbert was arguably the most influential figure during the inter-war years.

8.2. *The status of metalogic.* In 1918 metalogic was little recognised as independent of (object-level) logic. But the rise of metamathematics with Hilbert, the work on model theory by Skolem and others, and the study of problems in semantics (especially by the Poles) brought metalogic to the forefront of concern by the middle 1920's, where it has remained ever since.

8.3. *A logical revolution?* In 1918 logic was still an "occasional" interest, pursued by a few figures mostly working in isolation. Much of the writing in logic of that time was still discursive stuff written by philosophers for philosophers, with little symbolic content (although, in the work of the phenomenologist Husserl, much penetrating philosophical thought). But during the 1920's the subject was largely taken over by mathematicians, or at least by people with substantial mathematical training, and most major papers appeared in mathematical journals.

Thus, can we speak of a revolution in logic? The word 'revolution' is used so often by historians of science that I am disinclined to add myself to their voices; but we can certainly point to a substantial change of emphasis during this change of tradition. Among theories of scientific revolution, I am generally attracted to Kuhn's theory of normal science; but in this case the theory does not seem to function well. There was not really a 'normal' logic in the period before Frege, since various traditions were active and only the Boolean algebra made substantial use of mathematics. Nor was there a new paradigm afterwards, since the controversies between logicism, formalism, and intuitionism, already nascent in the birth of the latter two doctrines in the 1900's, prevented any of the three becoming a new 'normal' logic. Perhaps we can point to a change of paradigm in the much weaker, more general, sense of a conversion from general considerations, often involving psychology, to relatively specific concerns using mathematical techniques.

8.4. *Professional standing.* There were signs of professionalisation, especially in America and Poland: the Association for Symbolic Logic and its journal, *Studia philosophica*, and so on. In addition, quite a number of text-books were published, espousing one or another of the new developments in logic, and I believe that logic was being taught at university level rather more widely in some countries. However, I would not wish to exaggerate the extent of professionalisation; even today logic nestles under the wings of larger subjects. For example, there seems to have been little science policy applied to logic between the wars— though this was to change during the Second World War, when logic was applied to crypto-analysis and some logicians became code-breakers.

[6] A good impression of the three major philosophies of mathematics just prior to Gödel's 1931 theorem may be obtained from the papers by Carnap on logicism, Heyting on intuitionism, and von Neumann on formalism which were delivered to a Vienna Circle symposium on the foundations of mathematics in 1930. They were published in the Vienna Circle's journal *Erkenntnis* (1931), 91–121; English translations are contained in Benacerraf and Putnam *1964*, 31–54.

8.5. *Non-classical logics.* These logics did not progress very far between the wars; intuitionistic logic aroused the most attention. The interest in many-valued logics was rather slight (see Rutz *1973*); even their bearing on quantum mechanics was apparently not given prominence until Birkhoff and von Neumann *1936*. Modal logics seem also to have been rather unpopular (see Zeman *1972*), although they have their origins in Lewis *1918*.

8.6. *Interest in the history of logic.* Through the inter-war period interest in the history of logic, both ancient and relatively modern, maintained a steady low level. Of the major figures discussed above, Łukasiewicz showed the deepest historical concern and inspired among some of his compatriots an interest that has continued to this day. Church's bibliography *1936–38* of symbolic logic, which graced the early pages of the *Journal of symbolic logic*, is an important historical aid. Editions of the works of Hilbert and Peirce (and also of Cantor and Dedekind) were prepared during the early 1930's.

8.7. *Mathematicians and logic.* I mentioned in 8.3 that mathematicians largely "took over" logic in the 1920's. The main areas of contact between mathematics and logic were, of course, in the relationship between arithmetic, set theory, and predicate calculi. In addition, abstract algebra furnished appropriate structures for aspects of modal theory; and then there was lattice theory, which sprang to prominence in the 1930's after lying dormant in Schröder's algebraic logic among other places (see Mehrtens *1979*), and which furnished Birkhoff and von Neumann with the quantum logic that I mentioned in 8.5.

One could certainly add other topics to this list; but nevertheless the main impression is that contacts between logic and mathematics—especially "working" mathematics—were rather scattered between the wars (and, indeed, both before and after that period). While mathematicians emphasize the need for rigorous proofs and exact definitions, and take a *general* interest in problems of axiomatisation and proof, they are often reluctant to study these "logical" questions *explicitly*; even the most "exact" mathematics rarely meets logicians' standards (see Corcoran *1973*). A good example is provided by the French, who always produce a remarkable number of great mathematicians but who show little interest in logic. Herbrand, by far the greatest French logician between the wars, had difficulty in obtaining a panel to judge his doctoral thesis *1930*, and he published it in Poland.

The mathematician's attitude to logic is best understood by considering the position of set theory. It became clear between the wars that the attempt of logicism to embrace set theory within logic was forlorn, although the line of division between the two fields was not clearly or definitively drawn. Now mathematicians were interested in set theory much more than in logic; and even then they tended to draw on the set-*topological* aspects of the subject, as a source of definitions, techniques, and theorems, rather than the *axiomatic* set theory of Zermelo and others, where most of the contacts with logic were made. This situation is well exemplified by mathematical analysis, where set topology continued to play a significant role but axiomatic set theory and logic were largely absent—indeed, logic was much less prominent than it had been in the 1890's and 1900's, when Peano brought the "new" mathematical logic into analysis,[7] and Whitehead and Russell were working out the details of their logicist programme. Ziehen made a perceptive judgement and prediction at the beginning of our period (*1917*, 78), with which most inter-war mathematicians and logicians could agree: 'Set theory is no part of logic but its favoured-daughter discipline, from whose inspirations it has many more results to await'.

Acknowledgements. This paper is based on a lecture delivered to the conference on 'The restructuring of science between the two world wars' held in Florence and Rome in the summer of 1980. I am grateful to the organisers for allowing this version of the lecture to appear in the MONTHLY. For comments on a draft I express thanks to C. Lejewski, G. Moore, B. Smith, and a referee.

[7]At that time Peano's group made Italy the leading country for mathematical logic. But Italian interest between the wars was at a low level.

Bibliography

Collected or selected editions of works, where existent, are listed first under a catchword. Several of these editions contain introductions and commentaries which serve as further secondary literature.

A subscript indicates the edition of a work which is cited, and a dagger '†' indicates which edition of an item is cited *by page number* in the text. 'E' denotes 'English translation'. All editorial comments are enclosed within square brackets. Volume numbers of journals are given *in italics*. As mentioned in the introduction, the bibliography contains full names and dates of principal inter-war figures.

Ackermann, Wilhelm Friedrich (1896–1962) *see* Hilbert and Ackermann.

Benacerraf, P. and Putnam, H.

1964. Philosophy of mathematics. Selected readings, Oxford.

Bernays, Paul (1888–1977) *see* Hilbert and Bernays.

Birkhoff, G. and von Neumann, J.

1936. 'The logic of quantum mechanics', *Ann. maths.*, *37*, 823–843; von Neumann *Works*, vol. 4, 105–125.

Bowne, G. D.

1966. The philosophy of logic 1880–1908, The Hague.

Brouwer, Luitzen Egbertus Jan (1881–1966)

Works. Collected works, vol. 1 (ed. A. Heyting: 1975, Amsterdam).

1927a. 'Über Definitionsbereiche von Functionen', *Math. Ann.*, *97*, 60–75; *Works*, 390–405. E: van Heijenoort *1967*, 448–463.

1927b. 'Intuitionistische Betrachtungen über den Formalismus', *Proc. Kon. Akad. Wetens. Amsterdam*, *31*, 374–379; *Works*, 409–414. E: van Heijenoort *1967*, 490–492†.

Bunn, R.

1980. 'Developments in the foundations of mathematics, 1870–1910', in I. Grattan-Guinness (ed.), *From the calculus to set theory, 1630–1910. An introductory history* (1980, London), 220–255.

Carnap, Rudolf (1891–1970)

1928. Die logische Aufbau der Welt, Vienna. E: *1967*.

1929. Abriss der Logistik ..., Vienna.

1967. The logical structure of the world, London.

Cavaillès, J.

1937. Méthode axiomatique et formalisme, Paris.

1938. Remarques sur la formation de la théorie abstraite des ensembles, 2 vols., Paris.

Church, Alonzo (1903–)

1932–33. 'A set of postulates for the foundation of logic', *Ann. maths.*, (2)*33*, 346–366; (2)*34*, 839–864.

1936. 'An unsolvable problem of elementary number theory', *Amer. j. maths.*, *58*, 345–363; Davis *1965*, 88–107.

1936–38. 'A bibliography of symbolic logic', *J. symb. logic*, *1*, 121–218; *3*, 178–212; (1938, Princeton). [Addenda in later volumes.]

1952. Introduction to mathematical logic, Princeton.

Chwistek, Leon (1884–1944)

1924–25. 'The theory of constructive types. (Principles of logic and mathematics)', *Ann. Soc. Polon. Math.*, *2*, 9–48; *3*, 92–141.

Corcoran, J.

1973. 'Gaps between logical theory and mathematical practice', in M. Bunge (ed.), *The methodological unity of science* (1973, Dordrecht), 23–50.

Curry, H. B. and Feys, R.

1968. Combinatory logic, vol. 1, Amsterdam.

Dauben, J. W.

1979. Georg Cantor ..., Cambridge, Mass.

Davis, M.

1965. (Ed.) *The undecidable ...*, New York.

Dumitriu, A.

1978. History of logic, vol. 4, Tunbridge Wells.

Frege, Friedrich Ludwig Gottlob (1848–1925)

1925. 'Erkenntnisquellen der Mathematik ...', in *Nachgelassene Schriften* (ed. H. Hermes and others: 1969, Hamburg), 286–294. E: *Posthumous writings* (1979, Oxford), 267–274.

Gentzen, Gerhard (1909–1945)

Papers. Collected papers (ed. M. E. Szabo: 1969, Amsterdam).

1935. 'Untersuchungen über das logische Schliessen', *Math. Ztsch., 39*, 176–210, 405–431. E: *Papers*, 68–131.

1936. 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Math. Ann., 112*, 493–565. E: *Papers*, 132–213.

Gödel, Kurt (1906–1978)

1930. 'Die Vollständigkeit der Axiome des logischen Funktionkalkuls', *Monats. Math. Phys., 31*, 349–360. E: van Heijenoort *1967*, 582–591.

1931. 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandte Systeme. I', *Monats. Math. Phys., 38*, 173–198. Es: *On formally undecidable propositions*... (1962, Edinburgh), 35–72; Davis *1965*, 4–38; van Heijenoort *1967*, 596–616.

Goldfarb, W.

1979. 'Logic in the twenties: the nature of the quantifier', *J. symb. logic, 44*, 351–368.

Grattan-Guinness, I.

1975. 'Wiener on the logics of Russell and Schröder...', *Ann. sci., 32*, 103–132.

1977. *Dear Russell—dear Jourdain*, London.

1979a. 'On Russell's logicism and its influence, 1910–1930', in *Proceedings of the 3rd International Wittgenstein Symposium, 1978* (1979, Vienna), 275–280.

1979b. 'In memoriam Kurt Gödel:...', *Hist. math., 6*, 294–304.

1980. 'Georg Cantor's influence on Bertrand Russell', *Hist. phil. logic., 1*, 61–93.

Guillaume, M.

1978. 'Axiomatique et logique', in J. Dieudonné (ed.), *Abrégé d'histoire des mathématiques 1700–1900*, vol. 2 (1978, Paris), 315–430.

Herbrand, Jacques (1908–1931)

Works. Ecrits logiques (ed. J. van Heijenoort: 1968, Paris).

Writings. Logical writings (ed. W. Goldfarb: 1971, Dordrecht). [E of *Works*.]

1930. 'Recherches sur la théorie de la démonstration', *Prace Towarz. Nauk. Warzaw.*, Wyd. 3, no. 33; *Works*, 35–153. Es: *Writings*, 44–202†; van Heijenoort *1967*, 525–581 [part only].

1931. 'Sur le problème fondamental de la logique mathématique', *Spraw. Posied. Towarz. Nauk. Warsaw.*, Wyd. 3, *24*, 12–56; *Works*, 167–207. E: *Writings*, 215–271.

Hermes, H.

1966. 'Zur Geschichte der mathematischen Logik und Grundlagenforschung in den letzten fünfundsiebzig Jahren', *Jber. Dtsch. Math.-Ver., 68*, 75–96.

Heyting, Arend (1898–1980)

1934. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie,* Berlin.

Hilbert, David (1862–1943)

Papers. Gesammelte Abhandlungen, vol. 3 (1935, Berlin).

1918. 'Axiomatisches Denken', *Math. Ann., 78*, 405–415; *Papers*, 146–156. [Abstract; *L'ens. math.*, (1)*19* (1917), 330–331; *Actes Soc. Helv. Sci. Nat., 99* (1917), pt. 2, 129–130.] E: *Phil. math., 7* (1971), 1–12.

1922. 'Neubegründung der Mathematik', *Abh. Math. Sem. Hamburg Univ., 1*, 157–177; *Papers*, 157–177.

1923. 'Die logischen Grundlagen der Mathematik', *Math. Ann., 88*, 151–165; *Papers*, 178–191.

1926. 'Über das Unendliche', *Math. Ann., 95*, 161–190; *Grundlagen der Geometrie*$_7$ [and later editions]. Es: van Heijenoort *1967*, 367–392; Benacerraf and Putnam *1964*, 134–151 [part].

Hilbert, D. and Ackermann, W. F.

1928. *Grundzüge der theoretischen Logik*$_1$, Berlin. [$_2$1938.] E: *Principles of mathematical logic* (1950, New York).

Hilbert, D. and Bernays, P.

1934–39. *Grundlagen der Mathematik*, 2 vols., Berlin.

Jordan, Z. A.

1945. *The development of mathematical logic and of logical positivism in Poland between the two wars,* Oxford; McCall *1967*, 346–406 [first six sections only].

Jörgensen, J.

1931. *A treatise of formal logic,* 3 vols., Copenhagen and London.

Kleene, Stephen Cole (1909–)

1936. 'General recursive functions of natural numbers', *Math. Ann., 112*, 727–742; Davis *1965,* 236–253.

1952. *Introduction to metamathematics*, Amsterdam and Groningen.

Kneale, W. and M.

1962. *The development of logic,* Oxford.

Kuratowski, K.

1980. *A half-century of Polish mathematics*... Warsaw and Oxford.

Leśniewski, Stanisław (1886–1939)

Works. [An edition of Leśniewski's works is soon to be published by Martinus Nijhoff (The Hague).]

1929. 'Grundzüge eines neuen Systems der Grundlagen der Mathematik', *Fund. math., 14*, 1–81.

Lewis, Clarence Irving (1883-1964)

1918. A survey of symbolic logic, Berkeley. [1960 Dover reprint is incomplete.]

Löwenheim, Leopold (1878-1957)

1915. 'Über Möglichkeiten im Relativkalkül', *Math. Ann.*, *76*, 447-470. E: van Heijenoort *1967*, 228-251.

Luschei, Eugene C.

1962. The logical systems of Leśniewski, Amsterdam.

Łukasiewicz, Jan (1878-1956)

Works. Selected works (ed. L. Borkowski: 1970, Amsterdam and Warsaw).

Łukasiewicz, J. and Tarski, A.

1930. 'Untersuchungen über den Aussagenkalkül', *C. r. Soc. Sci. Lett. Varsovie, 23,* cl. 3, 30-50. E: Łukasiewicz *Works*, 131-152; Tarski *Papers*, 38-59.

Mangione, C.

1972-76. 'La logica del ventesimo secolo', in L. Geymonat (ed.), *Storia del pensiero scientifico*, vol. 6 (1972, n.p.), 469-682; vol. 7 (1976, n.p.), 299-433.

McCall, S.

1967. (Ed.) *Polish logic 1920-1939*, Oxford. [See review by W. A. Pogorzelski in *J. symb. logic*, *35* (1970), 442-446.]

Mehrtens, H.

1979. Die Entstehung der Verbandstheorie, Hildesheim.

Moore, G. H.

1980. 'Beyond first-order logic: the historical interplay between mathematical logic and axiomatic set theory', *Hist. phil. logic*, *1*, 95-137.

1981. Zermelo's axiom of choice: its origins, development, and influence. Berlin.

Mostowski, A.

1965. Thirty years of foundational studies ..., Helsinki; *Selected works*, vol. 1 (1980, Amsterdam), 1-176.

Quine, Willard van Orman (1908-)

1937. 'New foundations for mathematical logic', this MONTHLY, *44*, 70-80; *From a logical point of view*$_2$ (1961, New York), ch. 4.

1940. Mathematical logic$_1$, Cambridge, Mass. [$_2$1951.]

1969. Set theory and its logic$_2$, Cambridge, Mass.

Ramsey, Frank Plumpton (1903-1930)

1926. 'The foundations of mathematics', *Proc. London Math. Soc.*, (2) *25*, 338-384; *The foundations of mathematics...* (ed. R. B. Braithwaite: 1931, London), 1-61; *Foundations* (ed. D. H. Mellor: 1978, London and Henley), 152-212.

Reid, C.

1970. Hilbert, Berlin.

Rescher, N.

1969. Many-valued logic, New York.

Richards, J.

1980. 'Boole and Mill: differing perspectives on logical psychologism', *Hist. phil. logic*, *1*, 19-36.

Rickey, V. F.

1977. 'A survey of Leśniewski's logic', *Studia logica*, *36*, 407-426.

Russell, Bertrand Arthur William (1872-1970) *see* Whitehead and Russell.

Rutz, P.

1973. Zweiwertige und mehrwertige Logik. Ein Beitrag zur Geschichte und Einheit der Logik, Munich.

Sierpiński, W.

1918. 'L'axiome de M. Zermelo et son rôle dans la théorie des ensembles et l'analyse', *Bull. Acad. Sci. Cracovie, cl. sci. math. nat.*, (A), 97-152; *Oeuvres choisies*, vol. 2 (ed. S. Hartman and A. Schinzel: 1975, Warsaw), 208-255.

Skolem, Thoralf (1887-1962)

Works. Selected works in logic (ed. J. E. Fenstad: 1970, Oslo).

1920. 'Logisch-kombinatorische Untersuchungen...', *Videns. skr. I mat. naturv. kl.*, no. 4; *Works*, 103-136. Part E: van Heijenoort *1967*, 252-263.

1923a.'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922...* (1923, Helsinki), 217-232; *Works*, 137-152. E: van Heijenoort *1967*, 290-301.

1923b.'Begründung der elementaren Arithmetik durch die recurrierende Denkweise', *Videns. skr. I. mat. naturv. Kl.*, no. 6; *Works*, 153-188. E: van Heijenoort *1967*, 302-333.

Sobociński, B.

1949–50. 'L'analyse de l'antinomie Russellienne par Leśniewski', *Methodus*, *1*, 94–107, 220–228, 308–316; *2*, 237–257.

Styazhkin, N. I.

1970. From Leibniz to Peano: a concise history of mathematical logic, Cambridge, Mass.

Tarski, Alfred Tajtelbaum (1902–)

Papers. Logic, semantics, metamathematics (transl. J. H. Woodger: 1956, Oxford).

1923. 'O wyrazie pierwotnym logistyki', *Prz. Filoz.*, *26*, 68–89. E: *Papers*, 1–23. [French version: *Fund. math.*, *4* (1923), 196–200; *5* (1924), 59–74.]

1930. 'Fundamentale Begriffe der deduktiven Wissenschaften. I', *Monats. Math. Phys.*, *37*, 361–404. E: *Papers*, 60–109.

1933. Pojęcie prawdy w językach nauk dedukcyjnych, Warsaw. E: *Papers*, 152–278. [German version: *Stud. phil.*, *1* (1936), 261–405.]

1935. O logice matematycznej i metodzie dedukcyjnej, Warsaw. E: *Introduction to logic and the methodology of the deductive sciences*$_1$ (1941, New York). [German edition: *Einführung in die mathematische Logik und die Methodologie der Mathematik* (1937, Vienna).]

See also Łukasiewicz and Tarski.

Turing, Alan Mathison (1912–1954)

1936. 'On computable numbers, with an application to the Entscheidungsproblem', *Proc. London Math. Soc.*, (2)*42*, 230–265 [with corrections in (2)*43* (1937), 544–546]; Davis *1965*, 116–154.

van Heijenoort, J.

1967. (Ed.) *From Frege to Gödel. A source book in mathematical logic 1879–1931*, Cambridge, Mass.

von Neumann, John (1903–1957)

Works. Collected works (ed. A. H. Taub, 6 vols.: 1961–63, New York).

1927. 'Zur Hilbert'schen Beweistheorie', *Math. Ztsch.*, *26*, 1–46; *Works*, vol. 1, 256–300.

See also Birkhoff and von Neumann.

Weyl, Claude Hugo Hermann (1885–1955)

1927. Philosophie der Mathematik und Naturwissenschaften, Munich. E: *Philosophy of mathematics and natural science* (1952, Princeton).

Whitehead, Alfred North (1861–1947)

1934. 'Indication, classes, numbers, validation', *Mind*, n.s. *43*, 281–297, 543 [corrigenda]; *Essays in science and philosophy* (1948, London), 227–240.

Whitehead, A. N. and Russell, B. A. W.

PM. Principia mathematica (3 vols., $_1$1910–13, $_2$1925–27, Cambridge).

Wittgenstein, Ludwig (1889–1951)

1922. Tractatus logico-philosophicus, London. [Revised E: (1961, London†).]

Zeman, J. J.

1972. Modal logic, Oxford.

Zermelo, Ernst Friedrich Ferdinand (1871–1953)

1935. 'Grundlagen einer allgemeinen Theorie der mathematischen Satzsysteme. (Erste Mitteilung.)', *Fund. math.*, *25*, 136–146.

Ziehen, T.

1917. Das Verhältnis der Logik zur Mengenlehre, Berlin.

POLYGONS, CIRCULANT MATRICES, AND MOORE-PENROSE INVERSES

EDWARD T. WONG

Department of Mathematics, Oberlin College, Oberlin, OH 44074

Introduction. If P is the polygon with vertices $p_1, p_2, \ldots, p_n$ in order, then for any integer m,

---

The author received his Ph.D. from the University of Rochester under the direction of R. E. Johnson. After two years at Yale, he joined the faculty of Oberlin College, where he has been since 1957 except for visiting appointments at Kyoto, Tunghai, and Tsing Hua universities (Taiwan), and Fudan University, Shanghai. His research interests are in algebra, especially ring theory and linear algebra.

$1 \leqslant m \leqslant n$, we can always construct another polygon Q, the m-descendant of P, with vertices $q_1, q_2, \ldots, q_n$ defined by

$$q_1 = \frac{1}{m}(p_1 + \cdots + p_m),\ q_2 = \frac{1}{m}(p_2 + \cdots + p_{m+1}), \ldots, \text{ and } q_n = \frac{1}{m}(p_n + p_1 + \cdots + p_{m-1}).$$

But the converse is false, that is, not every polygon Q is the m-descendant of some polygon P. (This follows trivially for $m = 2$ since the 2-descendant of any quadrilateral is always a parallelogram.) For convenience, we call Q *m-special* if Q is the m-descendant of some P. Edward Kasner **[3]**, **[4]**, and three of his students discovered that a polygon Q is m-special if and only if starting from any vertex of Q the vector sum of every dth side of Q is zero, where d is the g.c.d. of m and n. (For example, a polygon Q with an even number of vertices $q_1, \ldots, q_{2s}$ in order is 2-special if and only if $(q_2 - q_1) + (q_4 - q_3) + \cdots + (q_{2s} - q_{2s-1}) = 0$.) They also showed that if Q is m-special then there exists a unique m-special P where Q is the m-descendant of P. In their paper **[1]**, Berlekamp, Gilbert, and Sinden also mentioned a condition for a polygon not being 2-special. But their main concern is the eventual convexity of 2-descendants of a polygon. In this paper, we shall obtain all of Kasner's results and more from a different approach and show that they are immediate consequences of elementary properties of a linear transformation induced by a *circulant matrix*. If $V = \mathbb{C}^n$, where $\mathbb{C}$ is the field of complex numbers, then we can identify each vector $P = (p_1, \ldots, p_n)$ in V with the polygon having vertices $p_1, \ldots, p_n$ in order, and conversely. We show that if we denote the m-descendant mapping by T, that is, if TP is the m-descendant of P (T is a linear transformation on V), then there exists a linear transformation B on V such that (i) $(BT)Q = (TB)Q$ for all Q, and (ii) for each Q, whether m-special or not, BQ is the unique m-special polygon for which the sum of the squares of the distances between the corresponding vertices of $T(BQ)$ and Q is minimal, i.e., $\|T(BQ) - Q\| \leqslant \|TP - Q\|$ for all P. The mapping B turns out to be the *Moore-Penrose* inverse of T. In addition, we show that the Moore-Penrose inverse of a circulant matrix is also circulant.

Polygons. As mentioned in the Introduction, we identify each vector $P = (p_1, \ldots, p_n)$ in $V = \mathbb{C}^n$ with the polygon having vertices $p_1, \ldots, p_n$ in order. The m-descendant Q of P can be written $Q = TP$ where T is the $n \times n$ matrix

$$T = \frac{1}{m}\begin{bmatrix} \overbrace{1 & 1 & \cdots & 1 & 1}^{m} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 & 1 & 1 & \cdots & 0 \\ & & & \vdots & & & & \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

We also use T to denote the linear transformation on V induced by the matrix T relative to the standard basis of V. Hence our first task is to characterize the elements in the range R_T of T.

Circulant Matrices. The matrix T is a matrix of very special type, a so-called *circulant matrix*. An $n \times n$ matrix A over a field F is circulant if the first row $(a_1 a_2 \ldots a_{n-1} a_n)$ is arbitrary, but the second row is $(a_n a_1 \ldots a_{n-2} a_{n-1}), \ldots,$ and the last row is $(a_2 a_3 \ldots a_n a_1)$. It is clear that such a matrix is completely determined by its first row, each of the other rows being obtained by a cyclical permutation of the preceding row. Interest in circulant matrices has increased recently; see Davis **[2]**.

The simplest nonzero circulant matrix other than the identity is the matrix C whose first row is $(010\ldots0)$. It is easy to verify that $C, C^2, \ldots, C^{n-1}, C^n = I$, that all are circulant, and that $A = a_1 I + a_2 C + \cdots + a_n C^{n-1}$ if A is the circulant matrix with first row $(a_1 a_2 \ldots a_n)$.

THEOREM 1. *An $n \times n$ matrix A over a field F is circulant if and only if $AC = CA$.*

Proof. If A is circulant then we have already seen that A commutes with C. Let $V = F^n$ and $\alpha = (1, 0, \ldots, 0)$. Since $\{\alpha, C\alpha, \ldots, C^{n-1}\alpha\}$ is a basis of V, for any linear transformation (matrix)

A of V we have $A\alpha = a_1\alpha + a_2(C\alpha) + \cdots + a_n(C^{n-1}\alpha)$. If $AC = CA$, then

$$\left[A - \left(a_1 + a_2C + \cdots + a_nC^{n-1}\right)\right](C^i\alpha) = C^i\left(A\alpha - \left(a_1 + a_2C + \cdots + a_nC^{n-1}\right)\alpha\right) = 0$$

for all $i = 1, \ldots, n$. This implies $A = a_1 + a_2C + \cdots + a_nC^{n-1}$ and hence A is circulant.

Since the characteristic polynomial of C is $x^n - 1$, if the scalars are complex numbers then C has n distinct eigenvalues $\lambda_k = e^{ik\theta}$ where $\theta = 2\pi/n$ and $k = 1, \ldots, n$. If we let $U_k = (1/\sqrt{n})(\lambda_k, \lambda_k^2, \ldots, \lambda_k^n)$ then $CU_k = \lambda_k U_k$. Thus U_k is an eigenvector of C associated with the eigenvalue of λ_k. Since

$$U_i \cdot U_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$\{U_i\}$ is an orthonormal basis of $V = \mathbb{C}^n$. For each $P \in V$, $P = \sum_{i=1}^n (P \cdot U_i)U_i$. If $A = a_1I + a_2C + \cdots + a_nC^{n-1}$, then $AU_k = (a_1 + a_2\lambda_k + \cdots + a_n\lambda_k^{n-1})U_k$. This shows U_k is also an eigenvector of A associated with the eigenvalue $\lambda_k^* = a_1 + a_2\lambda_k + \cdots + a_n\lambda_k^{n-1}$. Hence $AP = \sum_{i=1}^n (P \cdot U_i)\lambda_i^* U_i$. The null space N_A of A is spanned by those U_i for which the corresponding $\lambda_i^* = 0$, whereas the range R_A of A is spanned by U_j with corresponding $\lambda_j^* \neq 0$. Hence $V = R_A \oplus N_A$ and $R_A = N_A^{\perp}$. The restriction of A on R_A is one-to-one and onto. For any Q in V, $Q \in R_A$ if and only if $Q \cdot U_i = 0$ for all i where $\lambda_i^* = 0$.

Now let $T = (1/m)(I + C + \cdots + C^{m-1})$, so that $TP = Q$ is the m-descendant of P. The eigenvalue λ_k^* of T is given by $\lambda_k^* = (1/m)(1 + \lambda_k + \cdots + \lambda_k^{m-1})$ or

$$\lambda_k^* = \begin{cases} \dfrac{1}{m}\left(\dfrac{1 - \lambda_k^m}{1 - \lambda_k}\right), & 1 \leqslant k \leqslant n - 1, \\ 1, & k = n. \end{cases}$$

Therefore $\lambda_k^* = 0$ if and only if $\lambda_k^m = e^{ikm2\pi/n} = 1$ or, equivalently, if and only if $k(m/n)$ is an integer for $1 \leqslant k \leqslant n - 1$. Let d be the g.c.d. of m and n and $m = dt$, $n = ds$. Since t and s are relatively prime and $1 \leqslant k \leqslant n - 1$, $k(m/n) = k(t/s)$ is an integer if and only if $d > 1$ and $k = s, 2s, \ldots, (d-1)s$. Thus T is invertible if and only if m and n are relatively prime. If $d > 1$ then $\lambda_k^* = 0$ if and only if $k = s, 2s, \ldots, (d-1)s$. Since $s\theta = s(2\pi/n) = s(2\pi/ds) = 2\pi/d$, $e^{is\theta} = e^{i(2\pi/d)}$, if we let $c = e^{i(2\pi/d)}$ then $1 + c^j + \cdots + c^{(d-1)j} = 0$ and

$$U_{js} = \frac{1}{\sqrt{n}}(c^j, c^{2j}, \ldots, c^{(d-1)j}, 1, c^j, \ldots, 1, \ldots, c^j, \ldots, 1)$$

for all $1 \leqslant j \leqslant d - 1$. Thus the null space N_T of T is of dimension $d - 1$ and has $\{U_{js}\}$, $1 \leqslant j \leqslant d - 1$, as an orthonormal basis. For any $Q \in V$,

$$Q = (q_1, \ldots, q_d, q_{d+1}, \ldots, q_{2d}, \ldots, q_{(s-1)d+1}, \ldots, q_{sd})$$

$$Q \cdot U_{js} = \frac{1}{\sqrt{n}}\Big[(\bar{c})^j(q_1 + q_{d+1} + \cdots + q_{(s-1)d+1}) + \cdots$$
$$+ (\bar{c}^j)^{d-1}(q_{d-1} + \cdots + q_{sd-1}) + 1(q_d + \cdots + q_{sd})\Big].$$

Collecting our results we have the following theorem.

THEOREM 2. *Let $T = (I + C + \cdots + C^{m-1})/m$ be the $n \times n$ circulant matrix over the complex numbers and d be the* g.c.d. *of m and n where $n = ds$, $m = dt$. Then the linear transformation T in $V = \mathbb{C}^n$ induced by T has the following properties*:

1. *T is invertible if and only if m and n are relatively prime.*
2. *If $d > 1$, then the null space N_T of T has dimension $d - 1$ and has orthonormal basis $\{U_{js}\}$, where $U_{js} = (1/\sqrt{n})(c^j, c^{2j}, \ldots, c^{(d-1)j}, 1, c^j, \ldots, 1, \ldots, c^j, \ldots, 1)$ with $c = e^{i(2\pi/d)}$ and $1 \leqslant j \leqslant d - 1$.*
3. *If $d > 1$ then a vector Q is in the range R_T of T if and only if $Q \cdot U_{js} = 0$ for all $j = 1, \ldots, d - 1$.*

4. *For each* $Q \in R_T$ *there exists a unique* $P \in R_T$ *such that* $TP = Q$.

Since our problem was formulated originally in geometric terms, the contents of Theorem 2, in some way, should be translated back to geometry. The following theorem does just that.

THEOREM 3. *Let* m *and* n *be positive integers,* $m \leqslant n$, *and* d *be the* g.c.d. *of* m *and* n. *Then*

1. *Every polygon* Q *of* n *vertices is the* m*-descendant of a unique polygon* P *if and only if* m *and* n *are relatively prime.*
2. *Let* $N = \{P \mid \text{the } m\text{-descendant of } P \text{ is the origin}\}$. *Then a polygon* W *is in* N *if and only if* $W = (w_1, \ldots, w_d, w_1, \ldots, w_d, \ldots, w_1, \ldots, w_d)$ *and* $w_1 + \cdots + w_d = 0$.
3. *If* $d > 1$ *and* $n = ds$, $m = dt$, *then* $Q = (q_1, \ldots, q_d, \ldots, q_{(s-1)d}, \ldots, q_{sd})$ *is* m*-special if and only if*

$$q_1 + q_{d+1} + \cdots + q_{(s-1)d+1} = q_2 + q_{d+2} + \cdots + q_{(s-1)d+2} = \cdots = q_d + q_{2d} + \cdots + q_{sd}.$$

Or, equivalently, Q *is* m*-special if and only if*

$$(*) \qquad \begin{cases} (q_2 - q_1) + (q_{d+2} - q_{d+1}) + \cdots + (q_{(s-1)d+2} - q_{(s-1)d+1}) = 0 \\ (q_3 - q_2) + (q_{d+3} - q_{d+2}) + \cdots + (q_{(s-1)d+3} - q_{(s-1)d+2}) = 0 \\ \quad \vdots \\ (q_d - q_{d-1}) + (q_{2d} - q_{2d-1}) + \cdots + (q_{sd} - q_{sd-1}) = 0. \end{cases}$$

Condition (*) *states that:* Q *is* m*-special if and only if starting from any vertex of* Q, *the vector sum of every* d*th side of* Q *is zero* [**4**, Theorem III].

4. *For each* m*-special polygon* Q *there exists a unique* m*-special* P *such that* Q *is the* m*-descendant of* P [**4**, Theorem V].

Proof. Parts 1 and 4 are simply the geometric formulations of parts 1 and 4, respectively, in Theorem 2. To prove part 2, first note that N is the null space N_T of T. Next observe that each vector in the basis $\{U_{js}\}$ of N_T, as given in part 2 of Theorem 2, satisfies the condition stated for W. Then any linear combination of $\{U_{js}\}$ also satisfies this condition. Hence part 2 is proved.

For part 3, if Q satisfies (*), let

$$x = q_1 + q_{d+1} + \cdots + q_{(s-1)d+1} = \cdots = q_d + q_{2d} + \cdots + q_{sd}.$$

Then $Q \cdot U_{js} = (x/\sqrt{n})(1 + (\bar{c})^j + \cdots + (\bar{c})^{d-1}) = 0$ for all $1 \leqslant j \leqslant d - 1$. Hence Q is m-special. If Q is m-special and $P = (p_1, \ldots, p_n)$ where $TP = Q$, then

$$q_1 = \frac{1}{m}(p_1 + \cdots + p_{dt}),\ q_2 = \frac{1}{m}(p_2 + \cdots + p_{dt+1}), \ldots,\ q_n = \frac{1}{m}(p_n + \cdots + p_{dt-1}).$$

Thus $q_i - q_{i-1} = (p_{dt+i-1} - p_{i-1})/m$, so that

$$\sum_{k=1}^{s-1} (q_{kd+2} - q_{kd+1}) = \frac{1}{m} \sum_{k=0}^{s-1} (p_{(k+t)d+1} - p_{kd+1})$$

$$= \frac{1}{m}\left(\sum_{k=0}^{s-1} p_{(k+t)d+1} - \sum_{r=0}^{s-1} p_{rd+1} \right).$$

For any k, let $k + t = hs + r$, $0 \leqslant r \leqslant s - 1$. It follows that $p_{(k+t)d+1} = p_{(ks+r)d+1} = p_{rd+1}$, since $sd = n$. This shows $\Sigma_{k=0}^{s-1}(q_{kd+2} - q_{kd+1}) = 0$. By starting from other vertices of P, we obtain the other equations in (*) precisely the same way. This completes the proof of part 3 and hence of Theorem 3.

Moore-Penrose Inverses. We begin with two questions:

Question 1. Given a polygon Q, how near is Q to being m-special? I.e., is there an m-special polygon "closest" to Q? If so, is it unique, and how can it be found? Algebraically, given a vector $Q \in C^n$, does there exist a unique vector $\tilde{Q} \in R_T$ where $\|T\tilde{Q} - Q\| \leq \|TP - Q\|$ for all P?

Question 2. If we call the m-descendant mapping T a *shrinking* mapping, then does there exist an *expanding* mapping B with the properties

$$\text{(a) } (TB)Q = (BT)Q \quad \text{and} \quad \text{(b) } (TBT)Q = TQ \text{ and } (BTB)Q = BQ$$

for all polygons Q? (Notice that if B exists then BQ is m-special for all Q since $BQ = (BTB)Q = T(B^2Q)$.)

Question 1 leads us immediately to consider the orthogonal projection of Q on R_T whereas question 2 leads us to consider the *Moore-Penrose* inverse (generalized inverse) of the m-descendant mapping T.

Let V and W be finite dimensional inner product spaces and A be a linear transformation from V into W. A linear transformation B from W into V is called a *Moore-Penrose inverse* of A if (i) $ABA = A$, $BAB = B$ and (ii) $(AB)^* = AB$ and $(BA)^* = BA$.

It is known, [**6**], [**7**], that if A has a Moore-Penrose inverse B then it is unique and for each $y \in W$, $\|A(By) - y\| \leq \|Ax - y\|$ for all $x \in V$. In fact, let $V = N_A^{\perp} \oplus N_A$ and $W = R_A \oplus R_A^{\perp}$ where N_A and R_A are the null space and range of A, respectively. Then for each $y \in W$ there exist unique $x \in N_A^{\perp}$ and $y' \in R_A^{\perp}$ for which $y = Ax + y'$. The mapping B from W to V defined by $By = x$ is the Moore-Penrose inverse of A [**7**].

Since our m-descendant mapping T is induced by a circulant matrix, we expect its Moore-Penrose inverse to be special. It is, as the following theorem shows.

THEOREM 4. *The Moore-Penrose inverse B of a complex circulant matrix A is also circulant and hence A and B commute.*

Proof. Previously, we have seen that if A is the linear transformation induced by a circulant matrix A on $V = \mathbb{C}^n$, then $V = R_A \oplus N_A$ and $R_A = N_A^{\perp}$. The set of vectors $\{U_i\}$ where the corresponding $\lambda_i^* \neq 0$ is an orthonormal basis of R_A. If B is the Moore-Penrose inverse of A, then $BU_i = B(A((1/\lambda_i^*)U_i)) = (1/\lambda_i^*)U_i$ if $\lambda_i^* \neq 0$ and $BU_j = 0$ if $\lambda_j^* = 0$. It follows that $BC = CB$. By Theorem 1, the matrix of B is also circulant.

Now let B be the Moore-Penrose inverse of T. Then $BT = TB$, $B = BTB = T^2B$, and $T = TBT$. For each vector $Q \in C^n$, $Q = T\tilde{Q} + Q'$ where $\tilde{Q} \in N_T^{\perp} = R_T$, $Q' \in R_T^{\perp} = N_T$, $BQ = \tilde{Q}$ and $\|T(BQ) - Q\| \leq \|TP - Q\|$ for all P. In geometric language, for each polygon Q there exists a unique m-special polygon $\tilde{Q} = BQ$ which is "closer" to Q than any other m-special polygons. We have thus proved the following result which answers both questions posed at the beginning of this section.

THEOREM 5. *Let m and n be positive integers $m \leq n$, T be the m-descendant mapping on the set of polygons of n vertices, and B be the Moore-Penrose inverse of T. Then for each polygon Q we have* (i) $(BT)Q = (TB)Q$, (ii) *BQ is the unique m-special polygon having the property that $\|T(BQ) - Q\| \leq \|TP - Q\|$ for all polygons P. That is, the sum of the squares of the distances between the corresponding vertices of the m-descendant of BQ and of Q is minimal among all m-special polygons.*

Example 1. Let

$$T = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Then the Moore-Penrose inverse B of T is given by

$$B = \frac{1}{4}\begin{bmatrix} 3 & -1 & -1 & 3 \\ 3 & 3 & -1 & -1 \\ -1 & 3 & 3 & -1 \\ -1 & -1 & 3 & 3 \end{bmatrix}.$$

For any quadrilateral $Q = (q_1, q_2, q_3, q_4)$. Our results show that

$$BQ = \frac{1}{4}(3q_1 - q_2 - q_3 + 3q_4, 3q_1 + 3q_2 - q_3 - q_4, -q_1 + 3q_2 + 3q_3 - q_4, -q_1 - q_2 + 3q_3 + 3q_4)$$

is the unique parallelogram having the property that

$$\|T(BQ) - Q\| \leqslant \|TP - Q\|$$

for all quadrilaterals P.

Example 2. Let

$$T = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then

$$B = \frac{1}{12}\begin{bmatrix} 11 & -7 & 2 & -7 & 11 & 2 \\ 2 & 11 & -7 & 2 & -7 & 11 \\ 11 & 2 & 11 & -7 & 2 & -7 \\ -7 & 11 & 2 & 11 & -7 & 2 \\ 2 & -7 & 11 & 2 & 11 & -7 \\ -7 & 2 & -7 & 11 & 2 & 11 \end{bmatrix}$$

is the Moore-Penrose inverse of T. For any hexagon $P = (p_1, p_2, \dots, p_6)$ with vertices $p_1, \dots, p_6$ in order, TP is the so-called parhexagon (defined by E. Kasner [5]) obtained from P by joining the centroids of each three consecutive vertices of P in order. For any hexagon Q, BQ is the unique parhexagon having the property that

$$\|T(BQ) - Q\| \leqslant \|TP - Q\| \quad \text{for all } P.$$

REMARKS. One can define in the same way the Moore-Penrose inverse of a linear transformation A between *arbitrary* inner product spaces V and W. If B is a Moore-Penrose inverse of A, then one can prove it is unique and for each $y \in W$, $\|A(By) - y\| \leqslant \|Ax - y\|$ for all $x \in V$. A necessary and sufficient condition for the existence of B is that the null space and range of A are closed in V and W, respectively.

The following method is a rather effective way to calculate the Moore-Penrose inverse of a matrix, especially by using a computer. For an $m \times n$ matrix A, let A_1 and D be the row reduced forms of A and AA^*, respectively. Then let $H = A_1^*$ and $K = AH$. The $m \times m$ matrix $E = K + (I - D)$, where I is the $m \times m$ identity matrix, is invertible and the $n \times m$ matrix $B = HE^{-1}$ is the Moore-Penrose inverse of A.

References

1. E. R. Berlekamp, E. N. Gilbert, and F. W. Sinden, A polygon problem, this MONTHLY, 72 (1965) 233–241.
2. Philip J. Davis, Circulant Matrices, Wiley, New York, 1979.
3. Edward Kasner, The group generated by central symmetries, with application to polygons, this MONTHLY, 10 (1903) 57–63.

4. Edward Kasner, E. Jonas, K. Way, and G. Comenetz, Centroidal polygons and groups, Scripta Mathematica, 2 (1934) 131–138.

5. Edward Kasner and James Newman, Mathematics and the Imagination, Simon and Schuster, New York, 1940.

6. Gilbert Strang, Linear Algebra and Its Applications, Academic Press, New York, 1980.

7. Edward T. Wong, Generalized inverses as linear transformations, Math. Gaz., 63, no. 425 (October 1979) 176–181.

ALMOST-PERIODIC FUNCTIONS

ROGER L. COOKE

Department of Mathematics, University of Vermont, Burlington, VT 05405

1. Introduction. Harald Bohr's theory of almost-periodic functions has been called "elegant" and extensively studied, especially in connection with differential equations. It has generally been considered too specialized, however, to form part of a standard course in real analysis, or even harmonic analysis. To be sure, it provides an application of compact symmetric operators in the text of Riesz and Nagy [**9**, pp. 254–259]; and it is presented as an illustration of the theory of Banach algebras in Loomis's book on harmonic analysis [**8**, pp. 165–173]. Nevertheless the basic Fourier-type theorems on pointwise convergence of the Fourier series of an almost-periodic function are not generally presented; and this omission is unfortunate, since such a study can do much to broaden and unify the student's understanding of this part of analysis.

In the present article we present a brief development of the theory of Fourier series of almost-periodic functions in order to show just how this subject can improve one's perspective on classical analysis. The points we wish to make are the following:

1. An almost-periodic function may be regarded as something intermediate between a periodic function and a function integrable over the entire line. Like the former, it has a Fourier series which represents it in some sense. In some respects, however, its Fourier series resembles the Fourier transform of the latter. In fact a formula for a "partial sum" of the series can be given which is literally correct for inverting the Fourier transform as well as for summing the Fourier series. This unifying formula was pointed out by Bochner [**4**] for functions of several variables.

2. In the standard treatment of convergence and summability using approximate identities the role played by the positivity of the approximate identity is not very extensive, and sometimes great advantages can be gained by sacrificing positivity. This advantage is clearly seen in the paper by Bochner just referred to; Bochner used as an approximate identity the Fourier transform of the Bochner-Riesz kernel $(1-|x|^2)^\delta\chi_{(0,1)}(|x|)$. The transform of this function, which is essentially a power of $|x|$ times a Bessel function of $|x|$, is not positive; its indefinite integrals have very good asymptotic properties, however. Bochner exploited these asymptotic properties in a manner which we shall partly explain below in order to obtain some very strong convergence theorems.

(One notable fact which becomes apparent in the course of a study of almost-periodic functions cannot, unfortunately, be presented here. It is that the impression conveyed by most treatments of Fourier series—that the convergence of the series is enhanced by the smoothness of the function it represents—is somewhat misleading. Actually, the various hypotheses of differentiability, Lipschitz conditions, bounded variation, etc., which seem to enhance convergence, do so only because the Fourier exponents of a periodic function have no finite point of accumulation.

Roger Cooke did his graduate work at Princeton University under the direction of Salomon Bochner. From 1966 to 1968 he taught at Vanderbilt University; since 1968 he has been at the University of Vermont. His research interest since 1980 has been the history of mathematics and he is spending a sabbatical year during 1981–82 in Stockholm, Moscow, and London studying the works of S. V. Kovalevskaya.—*Editors*

For almost-periodic functions, whose Fourier exponents may cluster around a finite point such as $x = 0$, convergence of the Fourier series is equally likely to be enhanced when the function the series represents has bounded indefinite integrals. The reader is referred to the works of Bredihina in the bibliography for details and justification of these statements.)

It may help to put our development of this topic into perspective if we look at a small part of its history at this point. (More historical notes will be found after the presentation of the main results.) In his book on almost-periodic functions [1] Besicovitch proved that the criteria for the convergence of Fourier series of periodic functions apply to the Fourier series of almost-periodic functions, provided the functions are of two special types. First, functions whose Fourier exponents do not get arbitrarily close together; second, functions which are the uniform limit of purely periodic functions. Besicovitch remarked that generalization to a more or less general class of almost-periodic functions seemed to be extremely unlikely, and adduced as evidence for this assertion the fact that smoothness of almost-periodic functions did not in general seem to enhance convergence. When Besicovitch was writing, the structure an almost-periodic function must have in order to guarantee convergence of its Fourier series was not yet clear. In a series of about two dozen papers written between 1955 and 1970, E. A. Bredihina elucidated this structure, proved some very strong convergence theorems, and provided counterexamples to show the optimality of these theorems.

Our development of the topic of almost-periodic functions aims at proving some theorems of the type proved by Bredihina. Our approach, however, is somewhat different. We shall adapt, with simplifications, an argument used by Bochner in the 1936 paper referred to above. In this way we shall be able to give a self-contained and elementary treatment which proceeds fairly far into the subject.

2. Preliminary Results. A. *The Definition of Almost-Periodicity.* What is an almost-periodic function? A simple example is the function $f(x) = \sin x + \sin\sqrt{2}\,x$. This function is clearly not periodic; for a continuous periodic function attains its maximum value once in every period. For $f(x)$ this maximum value is obviously 2; yet there is no number x for which $f(x) = 2$. It follows that there is no number p such that $f(x+p) = f(x)$ for all x. On the other hand there are many choices of p which make this equation approximately true for all x. In fact it is easy to see that for all x

$$|f(x+p) - f(x)| \leqslant |1 - \cos p| + |1 - \cos\sqrt{2}\,p| + |\sin p| + |\sin\sqrt{2}\,p|.$$

Now, if $\varepsilon > 0$ is given, let m and n be any integers such that $|m - \sqrt{2}n| < \varepsilon/4\pi$. If we choose $p = 2n\pi$, which gives $\cos p = 1$, $\sin p = 0$, we have

$$|f(x+p) - f(x)| \leqslant |1 - \cos\sqrt{2}\,p| + |\sin\sqrt{2}\,p|.$$

But then $\sqrt{2}p = (\sqrt{2}\,n)\cdot 2\pi = (m+\alpha)\cdot 2\pi$, where $|\alpha| < \varepsilon/4\pi$. Hence $\cos\sqrt{2}p = \cos 2\pi\alpha$, and $\sin\sqrt{2}p = \sin 2\pi\alpha$; and so

$$|f(x+p) - f(x)| \leqslant 2\pi|\alpha| + 2\pi|\alpha| = 4\pi|\alpha| < \varepsilon,$$

since $|1 - \cos\theta| \leqslant |\theta|$ and $|\sin\theta| \leqslant |\theta|$ for all θ.

It is not difficult to verify that there are many choices of m and n which satisfy $|m - \sqrt{2}n| < \varepsilon/4\pi$. Hence there are many suitable numbers p, called ε-translates of f.

The reasoning just presented justifies applying the name "almost-periodic" to the function $f(x)$, or more generally to any finite linear combination of sines and cosines. A general almost-periodic function differs from such a function by arbitrarily small amounts. To be specific, a function defined on the real line with values in the complex numbers is called almost-periodic if it can be uniformly approximated with any desired degree of accuracy by such a finite linear

combination of sines and cosines, or, what is the same thing, by a finite polynomial $q(x)=\Sigma_\lambda c_\lambda e^{i\lambda x}$.

B. *The Uses of Almost-Periodic Functions.* Why are almost-periodic functions of interest? What are their uses? We shall give just two instances of their occurrence in differential equations, sufficient, we hope, to show their importance.

Example 1. Consider a homogeneous linear differential equation with constant coefficients

$$\frac{d^n y}{dx^n}+c_1\frac{d^{n-1}y}{dx^{n-1}}+\cdots+c_{n-1}\frac{dy}{dx}+c_n y=0.$$

It is well known that the general solution of this equation is

$$y=p_1(x)e^{\alpha x}+\cdots+p_r(x)e^{\theta x},$$

where $\alpha,\beta,\dots,\theta$ are the distinct complex roots of the algebraic equation

$$z^n+c_1z^{n-1}+\cdots+c_{n-1}z+c_n=0$$

and $p_j(x), j=1,2,\dots,r$, are polynomials of degree less than the multiplicity of the corresponding exponents $\alpha,\beta,\dots,\theta$.

A very natural question to ask is: Can this differential equation have solutions which are bounded on the entire line? The explicit form of the solution makes it clear that the answer is positive. In fact a solution of the form just given is bounded if and only if each polynomial $p_j(x)$ which is not identically zero is a constant and corresponds to a root of the algebraic equation which is purely imaginary. In other words the bounded solutions of this equation are precisely those of the form

$$y=a_1e^{i\lambda_1 x}+\cdots+a_ke^{i\lambda_k x},$$

for real numbers $\lambda_1,\lambda_2,\dots,\lambda_k$. In other words the bounded solutions of the equation are precisely those which are almost-periodic.

The example just given is the starting point for a vast amount of study of almost-periodic functions as solutions to both ordinary and partial differential equations. (A good reference for this work is the recent book of S. Zaidman [**11**]. Professor Zaidman is the originator of many of the basic theorems in this area.) A great deal of the work in this area amounts to proving that the bounded solutions to a differential equation are almost-periodic. If one has physical or mathematical reasons for believing that bounded solutions exist, the entire machinery of Fourier analysis for almost-periodic functions is then immediately available. For example, Bohr and Neugebauer proved that, if the forcing function $g(x)$ is almost-periodic, then all bounded solutions of the equation

$$\frac{d^n y}{dx^n}+c_1\frac{d^{n-1}y}{dx^{n-1}}+\cdots+c_{n-1}\frac{dy}{dx}+c_n y=g(x)$$

are almost-periodic.

Example 2. Bochner and von Neumann [**5**] studied the linear partial differential equation

$$\sum_{p=1}^{k}\frac{\partial}{\partial x_p}\left(\sum_{q=1}^{k}a_{pq}(x)\frac{\partial\Psi}{\partial x_q}\right)=\mu(x)\frac{\partial^2\Psi}{\partial t^2},$$

which contains the wave equation as a special case, by searching for solutions of the form

$$\Psi(x;t)=\Sigma\Psi_n(x)e^{i\beta_n t}.$$

This technique, especially in the case $k=1$ $(x=(x_1,\dots,x_k))$ amounts to the familiar separation-

of-variables technique. As Bochner and von Neumann pointed out, however, the explicit form of the solution assumed amounts to saying that the solution is an almost-periodic function of t. By regarding the functions $\Psi_n(x)$ as vectors in a certain function space, Bochner and von Neumann reduced part of the problem to showing that certain ordinary differential equations have almost-periodic solutions.

C. *The Analysis of the Almost-Periodic Function.* The definition we have chosen for almost-periodicity makes it seem likely that an almost-periodic function can be analyzed using Fourier series whose exponents need not be integers, and in fact may even be incommensurable with one another. In fact our definition says that an almost-periodic function is the uniform limit of a sequence of trigonometric polynomials

$$p^{(n)}(x) = a_{n1}e^{i\lambda_{n1}x} + \cdots + a_{nm_n}e^{i\lambda_{nm_n}x}.$$

The uniform convergence of $p^{(n)}(x)$ ought to entail the formal convergence of these coefficients and exponents to a series $\Sigma_\lambda c_\lambda e^{i\lambda x}$. This conjecture is accurate; however, several problems do arise. First, one is bound to ask if different sequences of trigonometric polynomials which converge uniformly to the same function might not converge formally to different series. If so, a single almost-periodic function would have more than one Fourier series. Second, after one establishes the existence and uniqueness of the Fourier series, one is faced with the problem of how to use it; basically this problem is the problem of showing that the series converges to the function it comes from. We thus have arrived at the basic problems of Fourier series for almost-periodic functions: (1) to prove that each such function has one and only one Fourier series; (2) to prove that the series represents the function (so that different functions cannot have the same series). The remainder of this section is devoted to establishing part (1). Part (2) is the main topic of Fourier analysis and will occupy the next section. In what follows, f and f_n denote bounded continuous complex-valued functions on the real line.

LEMMA 1 (Existence and Uniqueness of Fourier Series). *For each real number λ such that the limit*

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} f(x)e^{-i\lambda x}\,dx$$

exists, denote this number by $\hat{f}(\lambda)$.

(1) *If $\hat{f}_n(\lambda)$ is defined for λ in a subset D of the real line and for $n = 1, 2, \ldots$ and $f_n(x)$ converges uniformly to $f(x)$, then $\hat{f}(\lambda)$ is defined for all λ in D and $\hat{f}_n(\lambda)$ converges uniformly on D to $\hat{f}(\lambda)$.*

(2) *If $f(x)$ is almost-periodic, then $\hat{f}(\lambda)$ is defined for all real λ, and $\hat{f}(\lambda) = 0$ for all but a countable set of λ.*

Proof. Suppose $\lambda \in D$, and let $\varepsilon > 0$. Since $f_n(x)$ converges uniformly to $f(x)$, there exists N such that $|f_m(x) - f_n(x)| < \varepsilon$ for all $m > N$, $n > N$, and all x. We then clearly have $|\hat{f}_m(\lambda) - \hat{f}_n(\lambda)| \leqslant \varepsilon$ for all $m > N$, $n > N$, and all λ in D. It follows that $\hat{f}_n(\lambda)$ converges uniformly on D to a limit $\Lambda(\lambda)$. We then have

$$\left|\Lambda(\lambda) - \frac{1}{2T}\int_{-T}^{+T} f(x)e^{-i\lambda x}\,dx\right| \leqslant |\Lambda(\lambda) - \hat{f}_n(\lambda)|$$

$$+\left|\hat{f}_n(\lambda) - \frac{1}{2T}\int_{-T}^{+T} f_n(x)e^{-i\lambda x}\,dx\right| + \frac{1}{2T}\int_{-T}^{+T}|f_n(x) - f(x)|\,dx.$$

From what has been assumed and shown we may take n sufficiently large that $|f_n(x) - f(x)| < \varepsilon/3$ for all x and $|\Lambda(\lambda) - \hat{f}_n(\lambda)| < \varepsilon/3$. If we also choose T_0 so large that

$$\left|\hat{f}_n(\lambda) - \frac{1}{2T}\int_{-T}^{+T} f_n(x)e^{-i\lambda x}\,dx\right| < \frac{\varepsilon}{3}$$

for all $T > T_0$, we shall have

$$\left|\Lambda(\lambda) - \frac{1}{2T}\int_{-T}^{+T} f(x)e^{-i\lambda x}\,dx\right| < \varepsilon \quad \text{for} \quad T > T_0.$$

This proves first of all that $\hat{f}(\lambda)$ is defined for λ in D and that $\hat{f}(\lambda) = \Lambda(\lambda)$. Since $\hat{f}_n(\lambda)$ converges uniformly to $\Lambda(\lambda)$, it follows that $\hat{f}_n(\lambda)$ converges uniformly to $\hat{f}(\lambda)$ for all λ in D. Part (1) is now proved completely.

To prove part (2) we first note that, if $p(x)$ is a trigonometric polynomial, say,

$$p(x) = a_1 e^{i\lambda_1 x} + \cdots + a_k e^{i\lambda_k x},$$

it is easy to verify that $\hat{p}(\lambda_j) = a_j$ for $j = 1, 2, \ldots, k$ and that $\hat{p}(\lambda) = 0$ for all other values of λ. To prove part (2) it is sufficient to apply part (1) to a sequence of trigonometric polynomials converging to $f(x)$ uniformly on the entire real line. Lemma 1 is now proved completely.

We have now obtained from each almost-periodic function a unique trigonometric series $\Sigma_\lambda \hat{f}(\lambda)e^{i\lambda x}$ which we shall call the Fourier series of $f(x)$. Thus we have obtained a way of "analyzing" such a function, i.e., breaking it up into a "spectrum" $\{\hat{f}(\lambda)\colon \lambda$ a real number$\}$. This analysis is of limited usefulness, however, unless we also know how to "synthesize" the function from its spectrum. From the history of Fourier series of periodic functions one would naturally try to form a partial sum of the Fourier series of an almost-periodic function and then establish that these partial sums converge to the function the series represents. At this point some difficulties arise which do not occur for Fourier series of ordinary periodic functions. The natural definition of the partial sum of the Fourier series is

$$S_\omega(f; x) = \sum_{|\lambda| \leqslant \omega} \hat{f}(\lambda)e^{i\lambda x}.$$

Unfortunately this sum may contain an infinite number of terms; it seems, then, that we must look for partial sums to define our partial sums. This ambiguity was neatly removed by Bochner in his 1936 paper. Bochner showed that, although $S_\omega(f; x)$ may contain an infinite number of terms, it is nevertheless (often) the Fourier series of an almost-periodic function $f_\omega(x)$. Convergence of the original series may thus be taken to mean convergence of the functions $f_\omega(x)$; and no harm is done by this convention if the sum $S_\omega(f; x)$ contains only a finite number of terms, since in that case the function it represents is simply its sum in the ordinary sense. The convergence theorems we shall prove in Section 4 will be understood in this sense. (For almost-periodic functions which are the uniform limit of periodic functions another definition of partial sum, also due to Bochner, is natural but again ambiguous, cf. [**1**].)

3. Basic Lemmas. The present section contains a little additional notation and some special results which are needed to establish the fundamental convergence theorems of the next section. Throughout this section $f(x)$ denotes an almost-periodic function. We introduce the following notation:

(1) $A_0(x) = \hat{f}(0)$ and, for $\nu > 0$, $A_\nu(x) = \hat{f}(\nu)e^{i\nu x} + \hat{f}(-\nu)e^{-i\nu x}$. Thus the Fourier series of f can be written as $\Sigma_{\nu \geqslant 0} A_\nu(x)$.

(2) For $a \geqslant 0$, $h \geqslant 0$, $N_{a,h}$ is the number of exponents λ satisfying $\hat{f}(\lambda) \neq 0$ and $a < \lambda^2 \leqslant a + h$.

(3) $f_x(t) = (f(x + t) + f(x - t))/2$.

(4) $H(t) = (4/\pi t^2)((\sin t/t) - \cos t)$. (The function $H(t)$ is our replacement for the kernel $(\sin t/t)^2$ used by Besicovitch to establish the results on convergence in his book [**1**]. It is a special case of the Fourier transform of the Bochner-Riesz kernel referred to in the Introduction. Bochner derived many of the properties of this kernel by relying on formulas in the book by Watson [**10**]. Since for our purposes the kernel can be expressed in terms of elementary functions, it seems better to give a simpler derivation of these properties, thus making the article self-contained. The reader should note in particular that little is lost by sacrificing the positivity of the kernel used by

Besicovitch, and a great deal is gained in terms of the asymptotic behavior of the indefinite integrals of $H(t)$.)

LEMMA 2.

$$\int_0^\infty H(t)\cos at\,dt = \begin{cases} 1-a^2 & \text{if} \quad a^2 \leqslant 1 \\ 0 & \text{if} \quad a^2 \geqslant 1. \end{cases}$$

Proof. Let $F(z) = \int_0^\infty e^{-zt}t^{-2}(t^{-1}\sin t - \cos t)\,dt$. The function $F(z)$ is continuous in the closed half-plane $\operatorname{Re}(z) \geqslant 0$ and analytic in the open half-plane $\operatorname{Re}(z) > 0$. The integral we wish to evaluate is $4\pi^{-1}\operatorname{Re}(F(ia))$. It is easy to compute that for $\operatorname{Re}(z) > 0$,

$$\int_0^\infty e^{-zt}\cos t\,dt = z(z^2+1)^{-1} \quad \text{and} \quad \int_0^\infty e^{-zt}\sin t\,dt = (z^2+1)^{-1}.$$

(These last two integrals can be evaluated through integration by parts.) We shall use these two formulas to find the integral we need. For $\operatorname{Re}(z) > 0$, we have

$$F'(z) = -\int_0^\infty e^{-zt}t^{-1}(t^{-1}\sin t - \cos t)\,dt, \tag{1}$$

and

$$F''(z) = \int_0^\infty e^{-zt}(t^{-1}\sin t - \cos t)\,dt,$$

that is,

$$F''(z) = \phi(z) - z(z^2+1)^{-1} \tag{2}$$

where

$$\phi(z) = \int_0^\infty e^{-zt}t^{-1}\sin t\,dt,$$

so that

$$\phi'(z) = -(z^2+1)^{-1}.$$

(In particular $\phi(z) = -\arctan z + C$ if z is *real*. This fact doesn't help us, however, since we need $F(z)$ for *imaginary* values of z.)

If we integrate by parts in equation (1), taking $u = t^{-1}(t^{-1}\sin t - \cos t)$ and $dv = -e^{-zt}$, so that $du = (-2t^{-2}(t^{-1}\sin t - \cos t) + t^{-1}\sin t)\,dt$ and $v = z^{-1}e^{-zt}$, we find

$$F'(z) = 2z^{-1}F(z) - z^{-1}\phi(z). \tag{3}$$

Differentiating equation (3) gives

$$F''(z) = 2z^{-1}F'(z) - 2z^{-2}F(z) - z^{-1}\phi'(z) + z^{-2}\phi(z). \tag{4}$$

If we solve (3) for $F(z)$ and substitute the result in (4), we get

$$F''(z) = z^{-1}F'(z) + z^{-1}(z^2+1)^{-1}. \tag{5}$$

By (2), (3), and (5) we can express $F(z)$ in terms of $\phi(z)$ as

$$F(z) = \tfrac{1}{2}(z^2+1)\phi(z) - \tfrac{1}{2}z.$$

Since $F(z)$ is continuous in the closed right half-plane

$$4\pi^{-1}\operatorname{Re}(F(ia)) = \lim_{x\to 0+} 4\pi^{-1}\operatorname{Re}(F(x+ia)) = 2\pi^{-1}(1-a^2)\lim_{x\to 0+}\operatorname{Re}(\phi(x+ia)). \tag{6}$$

Now

$$\operatorname{Re}(\phi(x+ia)) = \int_0^\infty e^{-xt}(\cos at)t^{-1}\sin t\,dt$$
$$= \tfrac{1}{2}\int_0^\infty e^{-xt}t^{-1}\sin(1+a)t\,dt + \tfrac{1}{2}\int_0^\infty e^{-xt}t^{-1}\sin(1-a)t\,dt.$$

But it is well known (Problem 9 below) that

$$\lim_{x\to 0+}\int_0^\infty e^{-xt}t^{-1}\sin\beta t\,dt = \begin{cases} \tfrac{1}{2}\pi & \text{if } \beta > 0, \\ 0 & \text{if } \beta = 0, \\ -\tfrac{1}{2}\pi & \text{if } \beta < 0. \end{cases}$$

It follows that

$$\lim_{x\to 0+}\operatorname{Re}(\phi(x+ia)) = \begin{cases} \tfrac{1}{2}\pi & \text{if } a^2 < 1, \\ 0 & \text{if } a^2 > 1, \\ \tfrac{1}{4}\pi & \text{if } a^2 = 1. \end{cases} \tag{7}$$

Formulas (6) and (7) give the desired result.

LEMMA 3. *The functions $H(t)$, $H_1(t) = -\int_t^\infty H(s)\,ds$, and $H_2(t) = -\int_t^\infty H_1(s)\,ds$ are all $O(t^{-2})$ as $t\to\infty$; hence each is absolutely integrable on $(0,\infty)$.*

Proof. For $H(t)$ the assertion is obvious. If $g(s)$ is either $\sin s$ or $\cos s$, one can easily show through integration by parts that $g_1(t) = \int_t^\infty s^{-k}g(s)\,ds$ and $g_2(t) = \int_t^\infty g_1(s)\,ds$ are $O(t^{-k})$ as $t\to\infty$, provided k is positive. The estimates for H_1 and H_2 follow easily from these facts.

LEMMA 4. *Let $f(x)$ be an almost-periodic function with Fourier series $\sum_{\nu\geq 0}A_\nu(x)$. Then for each $\omega > 0$ the function*

$$T_\omega^f(x) = \int_0^\infty f_x(t\omega^{-1/2})H(t)\,dt$$

is also an almost-periodic function. Its Fourier series is

$$\sum_{\nu^2\leq\omega}(1-\nu^2\omega^{-1})A_\nu(x). \tag{8}$$

Proof. Suppose first that $f(x) = e^{i\lambda x}$, so that $f_x(t) = e^{i\lambda x}\cos\lambda t$. From Lemma 2

$$T_\omega^f(x) = \begin{cases} (1-\lambda^2\omega^{-1})e^{i\lambda x} & \text{if } \lambda^2 \leq \omega, \\ 0 & \text{if } \lambda^2 \geq \omega. \end{cases}$$

Whichever of these two relations holds between λ and ω, $T_\omega^f(x)$ is explicitly seen to be an almost-periodic function whose Fourier series is given by (8). Lemma 4 is thus proved for an exponential monomial. The lemma follows for trigonometric polynomials as an immediate corollary.

In the general case $f(x)$ is the uniform limit of a sequence of trigonometric polynomials $p^{(n)}(x)$. It follows that the averages $p_x^{(n)}(t)$ converge uniformly in x and t to $f_x(t)$. Since $H(t)$ is absolutely integrable over $(0,\infty)$, it follows that $T_\omega^{p^{(n)}}(x)$ converges uniformly in x and ω to $T_\omega^f(x)$. Thus the function $T_\omega^f(x)$ is almost-periodic. Since the Fourier series of f is the formal limit of the series of $p^{(n)}$ and a similar relation holds between the series of $T_\omega^f(x)$ and $T_\omega^{p^{(n)}}(x)$, the Fourier series of $T_\omega^f(x)$ is given by (8), as asserted. Lemma 4 is now proved completely.

5. Two Typical Convergence Theorems

THEOREM 1. *Let $f(x)$ be an almost-periodic function satisfying the following two conditions:*

(1) *$f''(x)$ is bounded and uniformly continuous (this condition is equivalent to the assumption that $f'(x)$ and $f''(x)$ are almost-periodic).*

(2) *For some* $\eta > 0$, $\alpha \geqslant 0$, $N_{\omega,\eta}$ *is finite for all* $\omega > \alpha$, *and* $N_{\omega,\eta} = O(\omega^2)$ *as* $\omega \to \infty$.

Then for all $\omega > \alpha$, *the partial sum* $\Sigma_{\nu \geqslant \omega} A_\nu(x)$ *of the Fourier series of* $f(x)$ *is the Fourier series of an almost-periodic function* $f_\omega(x)$ *and* $f_\omega(x)$ *tends uniformly to* $f(x)$ *as* $\omega \to \infty$.

Please bear in mind throughout the proof that η remains constant.

Proof. If we integrate by parts twice in the formula which defines $T^f_\omega(x)$, we find

$$T^f_\omega(x) = -f_x(0)H_1(0) + \omega^{-1/2} f'_x(0) H_2(0) + \omega^{-1}\int_0^\infty f''_x(\omega^{-1/2}t) H_2(t)\,dt.$$

Since $f_x(0) = f(x)$ and $H_1(0) = -1$, we have

$$T^f_\omega(x) - f(x) = h(x)\omega^{-1/2} + \omega^{-1}Q(\omega, x),$$

where $h(x) = H_2(0) f'_x(0)$ is bounded because $f(x)$ and $f''(x)$ are both bounded. Let B be a fixed positive number. Since $f''_x(\omega^{-1/2}t)$ is uniformly bounded in x, t, and ω and tends uniformly on the set $0 < t \leqslant B$, $-\infty < x < \infty$, to $f''_x(0)$ as $\omega \to \infty$, while $H_2(t)$ is absolutely integrable on $(0, \infty)$, it follows that $Q(\omega, x)$ tends uniformly in x to the limit $f''_x(0)\int_0^\infty H_2(t)\,dt$.

We now rewrite the last equation in the form

$$\omega T^f_\omega(x) - \omega f(x) = h(x)\omega^{1/2} + Q(\omega, x). \tag{9}$$

If we replace ω by $\omega + \eta$ in equation (9) and then subtract (9) from the result, we obtain

$$(\omega + \eta)T^f_{\omega+\eta}(x) - \omega T^f_\omega(x) - \eta f(x) = h(x)\left(\sqrt{\omega + \eta} - \sqrt{\omega}\right) + Q(\omega + \eta, x) - Q(\omega, x). \tag{10}$$

From what has just been established about $h(x)$ and $Q(\omega, x)$ the first term on the right-hand side of (10) is uniformly $O(\omega^{-1/2})$ while the difference of the two Q's is uniformly $o(1)$.

Thus the function $g_\omega(x) = \eta^{-1}((\omega + \eta)T^f_{\omega+\eta}(x) - \omega T^f_\omega(x))$ tends uniformly to $f(x)$. This function is almost-periodic and its Fourier series is

$$\sum_{\nu^2 \leqslant \omega} A_\nu(x) + \sum_{\omega < \nu^2 \leqslant \omega + \eta} \left(1 - \frac{\nu^2 - \omega}{\eta}\right) A_\nu(x).$$

Now for $\omega > \alpha$ the second sum in this expression is the finite trigonometric polynomial

$$p_\omega(x) = \sum_{\omega < \nu^2 \leqslant \omega + \eta} c_\nu \hat{f}(\nu) e^{i\nu x}$$

where $0 \leqslant c_\nu < 1$. Thus for $\omega > \alpha$ the partial sum $\Sigma_{\nu^2 \leqslant \omega} A_\nu(x)$ is the Fourier series of the almost-periodic function

$$f_\omega(x) = g_\omega(x) - p_\omega(x).$$

Since we have already noted that $g_\omega(x)$ tends uniformly to $f(x)$, the theorem will be proved if we show that $p_\omega(x)$ tends uniformly to 0. To show this note that

$$|p_\omega(x)| \leqslant \sum_{\omega < \nu^2 \leqslant \omega + \eta} |\hat{f}(\nu)| \leqslant \left(\sum_{\omega < \nu^2 \leqslant \omega + \eta} |\hat{f}(\nu)|^2\right)^{1/2} N^{1/2}_{\omega,\eta}.$$

Since f'' is almost-periodic, Bessel's inequality (Problem 4(a) below) implies that

$$\sum_{\omega < \nu^2 \leqslant \omega + \eta} |\hat{f}''(\nu)|^2 = o(1).$$

Since $|\hat{f}''(\nu)|^2 = \nu^4 |\hat{f}(\nu)|^2$, we obtain

$$\sum_{\omega<\nu^2\leqslant\omega+\eta} |\hat{f}(\nu)|^2 = o(\omega^{-2}).$$

Then the assumption $N_{\omega,\eta} = O(\omega^2)$ implies that $p_\omega(x)$ is uniformly $o(1)$. This proves the theorem.

THEOREM 2. *Let $f(x)$ be an almost-periodic function satisfying the following two conditions:*

(1) *$f'(x) \in \mathrm{Lip}(\beta)$ for some $\beta > \frac{1}{2}$.*

(2) *For some $\alpha \geqslant 0$ and $\eta > 0$, $N_{\omega,\eta}$ is finite for all $\omega > \alpha$ and $N_{\omega,\eta} = O(\omega)$ as $\omega \to \infty$.*

Then for all $\omega > \alpha$ $\Sigma_{\nu^2\leqslant\omega} A_\nu(x)$ is the Fourier series of an almost-periodic function $f_\omega(x)$, and $f_\omega(x)$ tends uniformly to $f(x)$ as $\omega \to \infty$.

Proof. In most respects the proof is similar to that of Theorem 1. We do a single integration by parts in the integral which defines $T_\omega^f(x)$ and get

$$T_\omega^f(x) - f(x) = \omega^{-1/2}\int_0^\infty f_x'(\omega^{-1/2}t)H_1(t)\,dt,$$

so that

$$(\omega+\eta)T_{\omega+\eta}^f(x) - \omega T_\omega^f(x) - \eta f(x) = \left(\sqrt{\omega+\eta} - \sqrt{\omega}\right)\int_0^\infty f_x'\left(t(\omega+\eta)^{-1/2}\right)H_1(t)\,dt$$
$$+ \sqrt{\omega}\int_0^\infty \left(f_x'\left(t(\omega+\eta)^{-1/2}\right) - f_x'(t\omega^{-1/2})\right)H_1(t)\,dt. \tag{11}$$

Our first hypothesis implies that $f'(x)$ is almost-periodic, hence bounded. Since $H_1(t)$ is absolutely integrable, the first term on the right-hand side of (11) is uniformly $O(\omega^{-1/2})$. As for the second term, break the integral into two parts, the first part being over the interval $(0, \omega^{1/2}\log\omega)$ and the second over $(\omega^{1/2}\log\omega, \infty)$. The integral over the first of these intervals will be uniformly $O(|(\omega+\eta)^{-1/2} - \omega^{-1/2})\omega^{1/2}\log\omega|^\beta)$, which is $O(\omega^{-\beta}(\log\omega)^\beta)$. The integral over the second interval will be $O(\omega^{-1/2}(\log\omega)^{-1})$ since $\int_t^\infty |H_1(s)|\,ds = O(t^{-1})$. Multiplying these estimates by the coefficient $\sqrt{\omega}$, we find that the second term on the right-hand side of (11) is $O(\omega^{1/2-\beta}(\log\omega)^\beta) + O((\log\omega)^{-1})$, which is $o(1)$ if $\beta > \frac{1}{2}$. Hence, as in Theorem 1, the function $g_\omega(x)$ tends uniformly to $f(x)$. The proof that the remainder $p_\omega(x)$ tends uniformly to 0 uses the assumption $N_{\omega,\eta} = O(\omega)$ and is similar to the proof given in Theorem 1. Theorem 2 is now proved.

REMARKS. Theorems 1 and 2 are of interest only when the function $f(x)$ has arbitrarily large Fourier exponents. (Otherwise $f_\omega(x) = f(x)$ for all large values of ω, and the convergence statements become trivial.) It is known [7, p. 67] that the function $f_\omega(x)$ can be extended to the entire plane as an entire function. It would be pleasant if all such functions had absolutely convergent Fourier series, so that the sum which defines $f_\omega(x)$ could really be regarded as a partial sum. If you look for conterexamples such as $\Sigma_{n=1}^\infty n^{-1}\sin 2^{-n}x$ you might be led to conjecture that such is the case. (The series just shown defines a function which is almost-periodic in a general sense, but not in the sense we have been discussing.) Bredihina, however, has shown [3] that such a function may have a Fourier series which, in a natural arrangement, diverges at every point.

As remarked above, our proof of Theorem 1 follows very closely the proof of a convergence theorem of Bochner for purely periodic functions of several variables, [4, Theorem IX, first half]. Bochner gave arguments for periodic functions of any number of variables by considering two cases according as the number of variables is odd or even. Theorem 1 follows the argument for an even number of variables. (The argument Bochner gave for an odd number of variables does not seem to generalize to almost-periodic functions, since it depends on the even spacing of the square-norms of the Fourier exponents.)

Lipschitz conditions, such as occur in Theorem 2, have been extensively studied. Such conditions, it turns out, are equivalent to best-uniform-approximation conditions. In fact if $E(\lambda)$ denotes the distance (in uniform norm) from $f(x)$ to the subspace of almost-periodic functions g such that $\hat{g}(\mu) = 0$ for $|\mu| > \lambda$, then $f(x) \in \operatorname{Lip}\beta$ if and only if $E(\lambda) = O(\lambda^{-\beta})$ as $\lambda \to \infty$, cf. [**2**].

5. Historical Notes. What we have taken as the definition of almost-periodic functions is actually a rather advanced characterization of this class of functions. The original definition of H. Bohr called a function $f(x)$ almost-periodic if it is continuous and for every $\varepsilon > 0$ the set of all ε-translates is relatively dense, meaning there is some length c such that every interval of length c contains an ε-translate. This definition is rather cumbersome, as can be seen by its expression in logical notation:

$$\forall \varepsilon \exists c \forall a \exists p\left[(a \leqslant p \leqslant a + c) \wedge \forall x |f(x+p) - f(x)| < \varepsilon\right].$$

Cumbersome though it is, this definition suffices to prove quickly that almost-periodic functions are bounded and uniformly continuous, and that this class is closed under constant multiples, squaring, and uniform limits. The seemingly simple operation of addition, however, presents a formidable obstacle. It is by no means trivial, using Bohr's definition, that the sum of two almost-periodic functions is almost-periodic.

Fortunately Bochner soon found a much simpler characterization of almost-periodic functions. He proved that a bounded continuous function is almost-periodic if and only if every sequence of real numbers $\{t_k\}_{k=1}^{\infty}$ contains a subsequence $\{t_{k_n}\}_{n=1}^{\infty}$ such that $f(t + t_{k_n})$ converges uniformly. With this characterization taken as a definition, all the elementary facts about the class are easy to prove. The equivalence of Bochner's definition with Bohr's is a consequence of Ascoli's theorem [**6**, pp. 283–284].

6. Problems

1. (a) Show that there is no x for which $\sin x + \sin\sqrt{2}\,x = 2$.
   (b) Find a sequence x_n such that $\sin x_n + \sin\sqrt{2}\,x_n$ tends to 2.
2. Suppose m and n are integers with $n > 0$ and $|m - \sqrt{2}n| < \varepsilon/4\pi$.

Prove that there is a second pair m', n' satisfying this same inequality with $n < n' < n + K$, where K is a positive constant not depending on m or n, but only on ε.

3. Show that an almost-periodic function is bounded and uniformly continuous.

Also show that the class of such functions is closed under addition, multiplication, translation, and uniform limits.

4. (a) For any two almost-periodic functions $f(x)$ and $g(x)$, show that the inner product

$$(f, g) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(t)\overline{g(t)}\,dt$$

is defined and has all the usual properties of an inner product, including the Schwarz inequality. Show that the monomials $e^{i\lambda x}$ form an orthonormal set, and deduce Bessel's inequality $\Sigma_\lambda |\hat{f}(\lambda)|^2 \leqslant (f, f)$.

(b) Could one similarly define a convolution for almost-periodic functions? If so, how would the Fourier series of the convolution be related to the series of the two factors?

5. Show that, if one bounded solution of the equation

$$\frac{d^n y}{dx^n} + c_1 \frac{d^{n-1} y}{dx^{n-1}} + \cdots + c_{n-1} \frac{dy}{dx} + c_n y = g(x)$$

is almost-periodic, then all bounded solutions are almost-periodic.

6. Show that, if $f(x)$ has period p, then

$$\hat{f}(\lambda)=\begin{cases} 0 & \text{if } \lambda \neq 2\pi n p^{-1} \text{ for all } n=0,\pm 1,\pm 2,\ldots, \\ \dfrac{1}{p}\displaystyle\int_0^p f(x)e^{-i\lambda x}\,dx & \text{if } \lambda = 2\pi n p^{-1} \end{cases}$$

Hint: Write

$$\int_{-Np}^{+Np} f(x)e^{-i\lambda x}\,dx=\int_0^p f(x)e^{-i\lambda x}\sum_{k=-N}^{N-1} e^{-i\lambda pk}\,dx.$$

The sum inside the integral does not depend on x and is bounded for all N, since it can be explicitly summed as a geometric series, provided $\lambda \neq 2\pi n p^{-1}$.

7. Show that $t^{3/2}H(t)$ is a Bessel function of the first kind.

8. Use the fact that $H(t)$ is absolutely integrable to show that $F(z)$ is continuous in the closed right half-plane and analytic in the open right half-plane.

9. Find the constant C in the equation $\phi(z)=-\arctan z+C$ for real z. Use this result to verify the values given for

$$\lim_{x\to 0+}\int_0^\infty e^{-xt}t^{-1}\sin\beta t\,dt.$$

(Consider what happens as z tends to ∞.)

10. Generalize Lemma 3 to all iterated integrals of $H(t)$.

11. If $f(x)$ is absolutely integrable over the entire real line, its Fourier transform is defined to be

$$\hat{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)e^{-ixy}\,dx.$$

When $\hat{f}(y)$ is also absolutely integrable, the Fourier inversion formula holds, and asserts that

$$f(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\hat{f}(y)e^{ixy}\,dy.$$

(a) Take $f(x)=1-x^2$ when $x^2<1$ and $f(x)=0$ when $x^2>1$. Use this inversion formula to derive Lemma 2.

(b) The function $T_\omega^f(x)$ can be formed for integrable f just as for almost-periodic f. What is the Fourier transform of $T_\omega^f(x)$ in terms of the Fourier transform of f?

12. Show that, if $f(x)$ is almost-periodic and $f'(x)$ is uniformly continuous, then $f'(x)$ is also almost-periodic. (Use the mean-value theorem to show that $f'(x)$ is the uniform limit of $n(f(x+n^{-1})-f(x))$. Then invoke Problem 3.)

Deduce that if $f(x)$ is almost-periodic and $f''(x)$ is bounded and uniformly continuous, then $f'(x)$ and $f''(x)$ are both almost-periodic.

13. Verify that $H_1(0)=-1$.

14. What is the difference between $(f')_x$ and $(f_x)'$? Between $(\hat{f})'$ and $\widehat{(f')}$? (In this article we have used only the latter of each pair.)

References

1. A. S. Besicovitch, Almost Periodic Functions, Dover, 1954.

2. E. A. Bredihina, Best-possible criteria for the convergence of Fourier series of almost-periodic functions, Soviet Math. Dokl., 11 (1970) no. 5.

3. ______, The divergence of Fourier series of almost-periodic functions, Soviet. Math. Dokl., 7 (1966) no. 6.

4. S. Bochner, Summation of multiple Fourier series by spherical means, Trans. Amer. Math. Soc., 40 (1936) 175–207.

5. S. Bochner and J. von Neumann, On compact solutions of operational differential equations I, Ann. of Math., 36 (1935).

6. Dunford and Schwartz, Linear Operators, Part I, Wiley, 1957.
7. A. M. Fink, Almost Periodic Differential Equations, Lecture Notes in Mathematics, no. 377, Springer, 1974.
8. L. H. Loomis, Abstract Harmonic Analysis, Van Nostrand, 1953.
9. F. Riesz and B. Sz. Nagy, Functional Analysis, Ungar, 1955.
10. G. N. Watson, Bessel Functions, 2nd ed., Cambridge, 1944.
11. S. Zaidman, Abstract Differential Equations, Pitman, 1979.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

DIFFERENTIATION OF ASYMPTOTIC FORMULAS

D. J. NEWMAN
Department of Mathematics, Temple University, Philadelphia, PA. 19122

The simple example $x + \sin x \sim x$ (as $x \to \infty$) shows that it is not valid to differentiate an asymptotic formula in general. On the other hand it is known, e.g., that $f(x) \sim x^2$ with $f(x)$ convex does imply that $f'(x) \sim 2x$, and one wonders to what extent this result can be generalized.

The most tempting generalization is that $f(x) \sim g(x)$ with f and g both convex and increasing yields $f' \sim g'$, but this last is simply not true! The right hypothesis on $g(x)$ seems instead to be that $1/g(x)$ be convex. This rather general condition contains most "garden variety" functions such as x^λ, $x \log x$, e^x, e^{x^2} (for $x \geqslant 1$), e^{e^x}, etc., and we therefore feel it worthwhile to record the result.

THEOREM. *Suppose that $f(x) \sim g(x)$ as $x \to \infty$, where f, g are C^1 functions. If $f(x)$ is increasing and convex while $1/g(x)$ is convex then $f'(x) \sim g'(x)$.*

Proof. First some preliminaries. Note that $f(x)$ must approach ∞ so that $g(x)$ must also. Thus $1/g(x) \to 0$ and so this *convex* function must be decreasing, which is to say that $g(x)$ is increasing.

Now fix ε, $0 < \varepsilon < \frac{1}{6}$, and let $0 < h < x/2$. Then for all large x we have

$$f(x-h) < (1+\varepsilon^2)g(x-h), \quad (1-\varepsilon^2)g(x) < f(x) < (1+\varepsilon^2)g(x),$$
$$f(x+h) < (1+\varepsilon^2)g(x+h). \tag{1}$$

Next the convexity of $f(x)$ yields

$$f(x+h) - f(x) \geqslant hf'(x) \geqslant f(x) - f(x-h). \tag{2}$$

Consequently, combining (1) and (2) gives

$$(1+\varepsilon^2)g(x+h) - (1-\varepsilon^2)g(x) > hf'(x) > (1-\varepsilon^2)g(x) - (1+\varepsilon^2)g(x-h). \tag{3}$$

Also, the convexity of $1/g(x)$ gives both

$$\frac{1}{g(x+h)} \geqslant \frac{1}{g(x)} + h\left(\frac{1}{g(x)}\right)' = \frac{g(x) - hg'(x)}{g^2(x)} \tag{4}$$

and

$$\frac{1}{g(x-h)} \geqslant \frac{1}{g(x)} - h\left(\frac{1}{g(x)}\right)' = \frac{g(x)+hg'(x)}{g^2(x)}. \tag{5}$$

Thus if we restrict $h < g/g'$ we may insert the bounds given by (4) and (5) into (3). We have, namely,

$$(1+\varepsilon^2)g(x+h) - (1-\varepsilon^2)g(x) - hg'(x) > h(f'(x) - g'(x))$$

so that

$$\frac{(1+\varepsilon^2)g^2}{g-hg'} - (1-\varepsilon^2)g - hg' > h(f'-g'),$$

or

$$\frac{h^2g'^2 + 2\varepsilon^2g^2 - hgg'\varepsilon^2}{g-hg'} > h(f'-g'),$$

so that

$$\frac{h^2g'^2 + 2\varepsilon^2g^2}{(g-hg')} > h(f'-g').$$

Similarly, we may obtain a lower bound, and the result is

$$\frac{h^2g'^2 + 2\varepsilon^2g^2}{h(g-hg')} > f'-g' > -\frac{h^2g'^2 + \varepsilon^2(2g^2 + hgg')}{hg}. \tag{6}$$

We may also use (4) in another way. Choose $h = x$ and use our hypotheses to conclude that

$$\frac{g(x)-xg'(x)}{g(x)} \leqslant \frac{g(x)}{g(2x)} \sim \frac{f(x)}{f(2x)} \sim \frac{f(x)}{f(2x)+f(0)} \leqslant \frac{1}{2}.$$

Hence, again for large x, we have

$$\frac{xg'(x)}{g(x)} > \frac{1}{3}. \tag{7}$$

But (7) tells us that the choice $h = \varepsilon g(x)/g'(x)$ is permissible in (6), and if we make this choice we get, finally,

$$\frac{3\varepsilon}{1-\varepsilon}g' > f'-g' > -(3\varepsilon + \varepsilon^2)g', \tag{8}$$

which, of course, completes the proof.

The work for this paper was supported in part by NSF grant #MCSFB-02171.

MISCELLANEA

58. I sometimes feel inclined to apply the historical method to the multiplication table. I should make a statistical inquiry among school-children, before their pristine wisdom has been biassed by teachers. I should put down their answers as to what 6 times 9 amounts to, I should work out the average of their answers to six places of decimals, and should then decide that, at the present stage of human development, this average is the value of 6 times 9.—P. E. B. Jourdain, *The Philosophy of Mr. B*rtr*nd R*ss*ll*, London, 1918, p. 88.

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis MO 63121.

GENERALIZING THE GENERALIZED MEAN-VALUE THEOREM

ALEXANDER ABIAN
Department of Mathematics, Iowa State University, Ames, IA 50011

In what follows all functions are real-valued functions of a real variable x, and $a \leqslant x \leqslant b$ where a and b are distinct real numbers.

Let f be a function having a derivative of order $n \geqslant 1$ at $x = a$. Let $(T_{n,a}f)(x)$ denote the nth Taylor polynomial of f at $x = a$, i.e.,

$$(T_{n,a}f)(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{1}$$

To facilitate our exposition, we have devised the following $n+2$ by $n+2$ determinant representation of $(T_{n,a}f)(x)$:

$$(T_{n,a}f)(x) = \frac{1}{1!2!\cdots n!}\begin{vmatrix} 0 & x^n & x^{n-1} & \cdots & x & 1 \\ f(a) & a^n & a^{n-1} & \cdots & a & 1 \\ f'(a) & na^{n-1} & (n-1)a^{n-2} & \cdots & 1 & 0 \\ f''(a) & n(n-1)a^{n-2} & (n-1)(n-2)a^{n-3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f^{(n)}(a) & n! & 0 & \cdots & 0 & 0 \end{vmatrix} \tag{2}$$

Similarly, the following $n+2$ by $n+2$ determinant representation is devised for the difference $f(x) - (T_{n,a}f)(x)$

$$f(x) - (T_{n,a}f)(x) = \frac{-1}{1!2!\cdots n!}\begin{vmatrix} f(x) & x^n & x^{n-1} & \cdots & x & 1 \\ f(a) & a^n & a^{n-1} & \cdots & a & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f^{(n)}(a) & n! & 0 & \cdots & 0 & 0 \end{vmatrix} \tag{3}$$

where the determinant appearing in (3) is obtained from that appearing in (2) by preceding it with a minus sign and by changing 0 in its $(1,1)$ entry to $f(x)$.

The above determinant representations of $(T_{n,a}f)(x)$ and $f(x) - (T_{n,a}f)(x)$, which may be interesting in their own rights, can be established without difficulty.

We prove below (using (3)) successive generalizations of the well-known Cauchy's Generalized Mean-Value Theorem [**1**, p. 108].

For the first generalization, we assume that f and g are functions with continuous first derivatives f' and g' for $a \leqslant x \leqslant b$ and having second derivatives f'' and g'' for $a < x < b$. Let us consider a function h_1 given by:

$$h_1(x) = f(x) + pg(x) + qx + r \tag{4}$$

where p, q, r are real numbers, and let us determine p, q, r in such a way that:

$$h_1(b) = h_1(a) = 0 \tag{5}$$

and

$$h_1'(a) = 0. \tag{6}$$

Thus, p, q, r must satisfy the following three equations:

$$\begin{aligned} f(b) + pg(b) + qb + r &= 0 \\ f(a) + pg(a) + qa + r &= 0 \\ f'(a) + pg'(a) + q &= 0 \end{aligned}$$

from which (in view of (3)) we obtain

$$p = -\frac{\begin{vmatrix} f(b) & b & 1 \\ f(a) & a & 1 \\ f'(a) & 1 & 0 \end{vmatrix}}{\begin{vmatrix} g(b) & b & 1 \\ g(a) & a & 1 \\ g'(a) & 1 & 0 \end{vmatrix}} = -\frac{f(b) - (T_{3,a}f)(b)}{g(b) - (T_{3,a}g)(b)} \tag{7}$$

provided the denominators in (7) do not vanish.

From (4), in view of our hypothesis, it follows that h_1 has a continuous first derivative h_1' for $a \leqslant x \leqslant b$ and has a second derivative h'' for $a < x < b$. In fact,

$$h''(x) = f''(x) + pg''(x). \tag{8}$$

From (5), in view of Rolle's theorem, it follows that

$$h_1'(c) = 0 \quad \textit{for some } c \textit{ with } a < c < b. \tag{9}$$

But then again from (6) and (9), in view of Rolle's theorem, it follows that for some e with $a < e < c$, and therefore with $a < e < b$, we have $h_1''(e) = 0$, which by (8) implies:

$$f''(e) + pg''(e) = 0. \tag{10}$$

Assuming that f'' and g'' do not vanish simultaneously, from (10) we have

$$g''(e) \neq 0. \tag{11}$$

Substituting p given by (7) in (10), in view of (11) and (1) we obtain:

$$\frac{f(b) - (T_{3,a}f)(b)}{g(b) - (T_{3,a}f)(b)} = \frac{f''(e)}{g''(e)} = \frac{f(b) - f(a) - (b-a)f'(a)}{g(b) - g(a) - (b-a)g'(a)} \quad \textit{for some } e \textit{ with } a < e < b.$$

From the above it is clear (employing notation (3)) how to state and how to prove the following generalization of the generalized Mean-Value theorem.

THEOREM. *Let $n \geqslant 1$ be a natural number and let f and g be functions with continuous nth derivatives for $a \leqslant x \leqslant b$ and having $n+1$th derivatives $f^{(n+1)}$ and $g^{(n+1)}$ for $a < x < b$. Then*

$$\frac{f(b) - (T_{n,a}f)(b)}{g(b) - (T_{n,a}g)(b)} = \frac{f^{(n+1)}(u)}{g^{(n+1)}(u)} \quad \textit{for some } u \textit{ with } a < u < b \tag{12}$$

provided the denominator on the left-hand side of equality (12) *does not vanish and provided $f^{(n+1)}$ and $g^{(n+1)}$ do not vanish simultaneously.*

Proof. We consider a function h_n given by:

$$h_n(x) = f(x) + pg(x) + qx^n + rx^{n-1} + \cdots + t \tag{13}$$

where $p, q, r, \ldots, t$ are $n+2$ real numbers. We determine these $n+2$ real numbers in such a way that:

$$h_n(b) = h_n(a) = h_n'(a) = \cdots = h_n^{(n)}(a) = 0. \tag{14}$$

From (13) and (14) we obtain p to be equal to the negative of the left-hand side of the equality appearing in (12). On the other hand, repeated application of Rolle's theorem to the functions appearing in (14) implies that $h_n^{(n+1)}(u) = 0$ for some u with $a < u < b$, which, in its turn, in view of (13) implies $f^{(n+1)}(u) + pg^{(n+1)}(u) = 0$. From this we obtain p to be equal to the negative of the right side of the equality appearing in (12). Thus, the theorem is proved.

REMARK. The following easy proof of a generalized L'Hospital's rule follows immediately from our theorem. Let $f(a) = g(a) = 0$, and let $n+1$ be the smallest natural number for which it is no longer true that $f^{(n+1)}(a) = 0 = g^{(n+1)}(a)$. Then from (12) it follows that

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{f^{(n+1)}(a)}{g^{(n+1)}(a)}$$

where we include the case when both sides are infinite, and where we assume that the $(n+1)$th derivatives of f and g are continuous at a.

Reference

1. A. Taylor, Advanced Calculus, Ginn, Boston, Mass., 1955.

ON THE CONVERGENCE OF HALLEY'S METHOD

G. ALEFELD
Fachbereich Mathematik, Technische Universität Berlin
Strasse des 17. Juni 135, 1 Berlin 12, West Germany

1. Introduction. A number of papers have been written about Halley's method, a third-order method for the solution of a nonlinear equation. (See, for example, [**8**].) For real-valued functions, this method is usually written as

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k) - \frac{1}{2}f''(x_k)\dfrac{f(x_k)}{f'(x_k)}}, \qquad k \geqslant 0. \tag{0}$$

This method is also called the method of tangent hyperbolas, as in [**3**], because x_{k+1} as given by (0) is the intercept with the x-axis of a hyperbola that is osculatory to the curve $y = f(x)$ at $x = x_k$. Construction of the appropriate hyperbola, given $f(x_k)$, $f'(x_k)$, and $f''(x_k)$, is an interesting exercise.

Many of the authors writing on Halley's method have, in particular, been concerned with developing a convergence theorem similar to the so-called Newton-Kantorovich theorem. (See, for example, [**5**, p. 421 ff].) The most far-reaching results can undoubtedly be found in [**3**], where also a comprehensive list of references is given. Error bounds, also, are given in [**3**]. These reflect for the first time the correct order of the error.

In this note we add some new results on Halley's method for real functions. From a remark by G. H. Brown, Jr. [**2**], Halley's method can be derived by applying Newton's method to the function $g(x) = f(x)/\sqrt{f'(x)}$. The adaptation of Theorem 7.1 of Ostrowski [**7**] to Halley's method gives us results on the existence and uniqueness of a zero and on the convergence to this

zero. In particular, we get error bounds that reflect correctly the order of the error. This method is demonstrated by a simple well-tried numerical example.

2. Theoretical Results.

THEOREM. *Let $f(x)$ be a real-valued function of the real variable x, and let $f(x_0)f'(x_0) \neq 0$ for some x_0. Furthermore let*

$$f'(x_0) - \tfrac{1}{2}f''(x_0)\frac{f(x_0)}{f'(x_0)} \neq 0.$$

Define

$$h_0 = -\frac{f(x_0)}{f'(x_0) - \tfrac{1}{2}f''(x_0)\dfrac{f(x_0)}{f'(x_0)}}, \qquad x_1 = x_0 + h_0,$$

and set

$$J_0 = \begin{cases} [x_0, x_0 + 2h_0], & h_0 > 0 \\ [x_0 + 2h_0, x_0], & h_0 < 0. \end{cases}$$

For $x \in J_0$ let f have a continuous third derivative. Suppose that f' doesn't change sign in J_0 and that with

$$g(x) = \frac{f(x)}{\sqrt{f'(x)}}$$

we have

$$|g''(x)| \leqslant M_0 \tag{1}$$

and

$$2|h_0|M_0 \leqslant |g'(x_0)|. \tag{2}$$

Then starting with x_0 the feasibility of Halley's method is guaranteed. All x_k are contained in J_0, and the sequence $\{x_k\}$ converges to a zero x^* of f (which is unique in J_0).

Defining

$$h_k = -\frac{f(x_k)}{f'(x_k) - \tfrac{1}{2}f''(x_k)\dfrac{f(x_k)}{f'(x_k)}}, \qquad k \geqslant 0,$$

$$J_k = \begin{cases} [x_k, x_k + 2h_k], & h_k > 0 \\ [x_k + 2h_k, x_k], & h_k < 0 \end{cases}, \qquad k \geqslant 0,$$

$$|g''(x)| \leqslant M_k, \qquad x \in J_k, \quad k \geqslant 0,$$

we have the error estimates

$$|x_{k+1} - x_k| \leqslant \tfrac{1}{2}\frac{M_{k-1}}{|g'(x_k)|}|x_k - x_{k-1}|^2, \qquad k \geqslant 1, \tag{3}$$

$$|x^* - x_{k+1}| \leqslant \tfrac{1}{2}\frac{M_{k-1}}{|g'(x_k)|}|x_k - x_{k-1}|^2, \qquad k \geqslant 1, \tag{4}$$

$$|x^* - x_k| \leqslant \frac{M_{k-1}}{|g'(x_k)|}|x_k - x_{k-1}|^2, \qquad k \geqslant 1. \tag{5}$$

If

$$|f'(x)| \leqslant N, \qquad x \in J_0, \tag{6}$$

and if

$$\frac{1}{2}\frac{1}{\sqrt{f'(x)}}\left|-\frac{f'''(x)}{f'(x)}+\frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2\right| \leqslant M, \qquad x \in J_0, \tag{7}$$

then we have the coarser estimates

$$|x_{k+1}-x_k| \leqslant \frac{NM}{|g'(x_k)|}|x_k-x_{k-1}|^3, \qquad k \geqslant 1, \tag{3'}$$

$$|x^*-x_{k+1}| \leqslant \frac{NM}{|g'(x_k)|}|x_k-x_{k-1}|^3, \qquad k \geqslant 1, \tag{4'}$$

$$|x^*-x_k| \leqslant \frac{2NM}{|g'(x_k)|}|x_k-x_{k-1}|^3, \qquad k \geqslant 1. \tag{5'}$$

Proof. In some essential parts the proof is similar to that of Theorem 7.1 in [**7**]. Because of $f'(x_0) \neq 0$, we can assume that $f'(x_0) > 0$. We then consider the function

$$g(x) = \frac{f(x)}{\sqrt{f'(x)}}, \qquad x \in J_0.$$

For $x \in J_0$, g has a continuous second derivative and we have

$$g'(x) = \sqrt{f'(x)} - \frac{1}{2}\frac{f''(x)f(x)}{\sqrt{f'(x)^3}},$$

from which

$$g''(x) = \frac{1}{2}g(x)\left[-\frac{f''(x)}{f'(x)}+\frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2\right]. \tag{8}$$

Furthermore

$$|g'(x)-g'(x_0)| = \left|\int_{x_0}^{x} g''(t)\,dt\right| \leqslant |x-x_0|M_0,$$

and, using (2), we have

$$|g'(x_1)-g'(x_0)| \leqslant |x_1-x_0|M_0 = |h_0|M_0 \leqslant \frac{|g'(x_0)|}{2}, \tag{9}$$

from which we may estimate

$$|g'(x_1)| \geqslant |g'(x_0)| - |g'(x_0)-g'(x_1)| \geqslant \frac{|g'(x_0)|}{2}. \tag{10}$$

Integrating by parts and using

$$h_0 = -\frac{g(x_0)}{g'(x_0)},$$

we get

$$\int_{x_0}^{x_1}(x_1-x)g''(x)\,dx = g(x_1)$$

and therefore

$$|g(x_1)| \leqslant \frac{1}{2}|h_0|^2 M_0. \tag{11}$$

Because of (10), the feasibility of Halley's method (0) is guaranteed for $k = 1$ and we have

$$x_2 = x_1 + h_1.$$

Since

$$h_1 = -\frac{g(x_1)}{g'(x_1)}$$

we have from (10) and (11)

$$|h_1| \leqslant \frac{|h_0|^2 M_0}{|g'(x_0)|}.$$

Finally in a way similar to that in [7]

$$2|h_1|M_0 \leqslant |g'(x_1)|, \tag{12}$$

and

$$|h_1| \leqslant \frac{1}{2}|h_0|.$$

From these inequalities it follows that x_2 lies in J_0 and J_1 is contained in J_0; that is, we have $M_1 \leqslant M_0$. Therefore (12) can be replaced by

$$2|h_1|M_1 \leqslant |g'(x_1)|. \tag{13}$$

For the sequence

$$x_{k+1} = x_k + h_k \tag{14}$$

as computed by Halley's method it holds in general that

$$h_k = -\frac{g(x_k)}{g'(x_k)}. \tag{15}$$

Therefore (13) shows that our assumptions remain true if we replace x_0 by x_1 and h_0 by h_1, and by x_k and h_k, respectively, in general. The convergence to a zero x^* that is unique in J_0 can now be proved as in [7]. To prove the error estimates we start with (11): We have

$$|g(x_1)| \leqslant \frac{1}{2}|h_0|^2 M_0$$

and, therefore, using (15),

$$|x_2 - x_1| = |h_1| = \left|\frac{g(x_1)}{g'(x_1)}\right| \leqslant \frac{M_0}{2|g'(x_1)|}|x_1 - x_0|^2.$$

This is (3) for $k = 1$. For $k > 1$ the assertion is proved by mathematical induction. We omit the details. Since $|x^* - x_{k+1}| \leqslant h_k$ we immediately get (4) from (3).

Finally we have

$$|x^* - x_k| \leqslant |x_{k+1} - x_k| + |x_{k+1} - x^*|,$$

and (5) is therefore proved by using (3) and (4). Applying the mean-value theorem we have, for $x \in J_{k-1}$,

$$|g''(x)| = \frac{1}{2}\left|\frac{f(x)}{\sqrt{f'(x)}}\left[-\frac{f'''(x)}{f'(x)} + \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2\right]\right| \leqslant NM|x - x^*| \leqslant 2NM|x_k - x_{k-1}|.$$

Replacing, therefore, in (3), (4), and (5) M_{k-1} by the upper bound $2NM|x_k - x_{k-1}|$, we establish (3′), (4′) and (5′). □

Without going into details of the proof we remark that the estimation (5) can further be improved. If we define

$$t_k = \frac{1}{|g'(x_k)|}|h_k|M_k, \qquad k \geq 0,$$

then

$$|x^* - x_k| \leq \frac{1}{1+\sqrt{1-2t_k}}\frac{M_{k-1}}{|g'(x_k)|}|x_k - x_{k-1}|^2, \qquad k \geq 1. \tag{5''}$$

This last can be proved in exactly the same manner as the corresponding inequality for Newton's method (see [**4**, p. 34 ff]).

Since one can easily show that

$$t_k \leq \frac{1}{2}\frac{t_{k-1}}{1-t_{k-1}}, \qquad k \geq 1,$$

(5″) can be replaced by

$$|x^* - x_k| \leq \frac{1}{1+\sqrt{1-\dfrac{t_{k-1}}{1-t_{k-1}}}}\frac{M_{k-1}}{|g'(x_k)|}|x_k - x_{k-1}|^2. \tag{5'''}$$

Since $t_k \to 0$, (5″) and (5‴) are asymptotically better than (5) by a factor of $\frac{1}{2}$.

3. Numerical Example. In order to compare our theorem with other results, we consider the simple well-tried example (see, for example, [**3**, p. 453, Tabelle 2])

$$f(x) = x^3 - 10.$$

We have

$$f'(x) = 3x^2, \quad f''(x) = 6x, \quad f'''(x) = 6.$$

As in [**3**], we choose $x_0 = 2$ and get

$$h_0 = -\frac{g(x_0)}{g'(x_0)} = 2.153\,846\,154,^{\dagger}$$

$$x_1 = x_0 + h_0 = 2.153\,846\,154, \quad \text{and} \quad J_0 = [2, 2.307\,692\,308].$$

For $x \in J_0$ we have

$$g''(x) \leq M_0 = 0.288\,675\,135.$$

Furthermore

$$|g'(x_0)| = 3.752\,776\,750;$$

hence the main inequality (2) holds.

We have

$$|g'(x_1)| = 3.731\,590\,624.$$

Using this value in (5) for $k = 1$ we have

$$|x^* - x_1| \leq \frac{M_0}{|g'(x_1)|}|x_1 - x_0|^2 = 0.001\,831\,001.$$

† All numerical values have been computed using a HP21 pocket calculator.

The actual error is

$$|x^* - x_1| \approx 0.000\,588\,556.^\dagger$$

The error estimation reflects the order of the actual error and is only three times as large as the actual error.

For h_1 we get

$$h_1 = 0.000\,588\,536.$$

Therefore we have

$$J_1 = [x_1, x_1 + 2h_1]$$
$$= [2.153\,846\,154, 2.155\,023\,227],$$

and, for $x \in J_1$,

$$g''(x) \in [-0.000\,946\,821, 0.000\,945\,788],$$

hence

$$M_1 = 0.000\,946\,821.$$

For x_2 we get the value

$$x_2 = 2.154\,434\,690,$$

and therefore

$$|g'(x_2)| = 3.731\,530\,346.$$

Using (5) for $k = 2$, we finally have

$$|x^* - x_2| \leq \frac{M_1}{|g'(x_k)|}|x_2 - x_1|^2 \approx 8.78 \times 10^{-11}.$$

The actual error for x_2 is

$$|x^* - x_2| \approx 2.93 \times 10^{-11}.$$

Therefore the actual error is overestimated only by a factor of 3. (5‴) gives us the estimate

$$|x^* - x_2| \approx 4.39 \times 10^{-11}.$$

None of the error estimates that are to be found in Tabelle 2 in [**3**, p. 453] give a smaller bound for the error than (5‴).

4. Conclusion. If we compare the assumptions of our theorem with other results (especially with those given in [**3**]), then it seems that the most drastic assumption is the requirement that f' doesn't change sign in J_0. We need this assumption, however, in order to define $g(x)$ in the whole interval J_0.

If one wants to use the more precise error estimate (5), then one has to compute the bound M_{k-1} for the second derivative of g. There is no essential difficulty in doing this since one can get these bounds very simply by using interval arithmetic in (8). See, for example, [**1**, p. 28 ff], or [**6**, p. 161 ff], or [**9**]. The same is true for the inequality (5′) and the bounds M and N appearing there.

† The "≈" sign means here and in the sequel that the number is rounded upwards in the usual manner.

In conclusion we remark that the most important applications of Halley's method are to nonlinear equations in Banach spaces. See, for instance, the discussion and examples in [3].

The generalization of the results of this paper to this general case and some numerical examples will be discussed in another paper.

5. **Acknowledgment.** The author wants to express his thanks to Professor Jon Rokne for reading the manuscript. Thanks also to the Department of Mathematics and to the Department of Computer Science of the University of Calgary, Alberta, for the hospitality during the academic year 1979/80 when this paper was written.

References

1. G. Alefeld and J. Herzberger, Einführung in die Intervallrechnung, Bibliographisches Institut, Mannheim, 1974.
2. G. H. Brown, Jr., On Halley's variation of Newton's method, this MONTHLY, 84 (1977) 726–727.
3. B. Döring, Einige Sätze über das Verfahren der tangierenden Hyperbeln in Banach-Räumen, Aplikace matematiky, 15 (1970) 418–464.
4. B. Döring, Über das Newtonsche Näherungsverfahren, Math.-Phys. Sem.-Ber., 16 (1963) 27–40.
5. J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York and London, 1970.
6. L. Rall, Computational Solution of Nonlinear Operator Equations, Wiley, New York, 1969.
7. A. M. Ostrowski, Solutions of Equations in Euclidean and Banach Spaces, Academic Press, New York and London, 1973.
8. J. F. Traub, Iterative Methods for the Solution of Equations, Prentice-Hall, Englewood Cliffs, N.J., 1964.
9. R. E. Moore, Methods and Applications of Interval Analysis, SIAM Publications, Philadelphia, 1979.

PROBLEMS AND SOLUTIONS

EDITED BY VLADIMIR DROBOT

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

***Solutions** should be sent to the addresses given at the head of each problem set.*

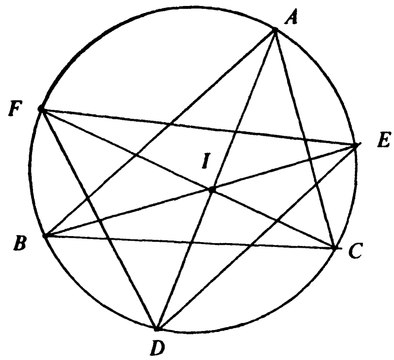
A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Variation on the Erdős-Mordell Geometric Inequality

S23 [1979, 863]. *Proposed by Jack Garfunkel, Flushing, N.Y., and Leon Bankoff, Los Angeles, Calif.*

Prove that the sum of the distances from the incenter of a triangle ABC to the vertices does not exceed half of the sum of the internal angle bisectors, each extended to its intersection with the circumcircle of triangle ABC. (See the figure.)



Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology, Eindhoven, the Netherlands. We use the following well-known inequality: if P is a point in the interior of $\triangle A_1A_2A_3$ at distances R_k and r_k from A_k and the side opposite A_k, respectively $(k = 1, 2, 3)$, then $R_1 + R_2 + R_3 \geqslant 2(r_1 + r_2 + r_3)$. (P. Erdős-L. J. Mordell, Problem 3740, this MONTHLY, 42 (1935) 396 and 44 (1937) 252–254).

Since $\overline{DB} = \overline{DI} = \overline{DC}$, $\overline{EC} = \overline{EI} = \overline{EA}$, $\overline{FA} = \overline{FI} = \overline{FB}$, the lines EF, FD, and DE are the perpendicular bisectors of $\overline{AI}$, $\overline{BI}$, and $\overline{CI}$, respectively. Hence I is the orthocenter of the triangle DEF and lies in the interior of this triangle. Therefore: $ID + IE + IF \geqslant IA + IB + IC$, which is equivalent to the proposed inequality.

Also solved by Steve Belbas, Jordi Dou (Spain), Leonard Goldstone, Hans Kappus (Switzerland), Naoki Kimura & Tetsundo Sekiguchi, L. Kuipers (Switzerland), V. N. Murty, C. R. Pranesacher (India), Achilles Venetoulias, and the proposers.

Notes: Most of the solvers pointed out that equality holds iff $\triangle ABC$ is equilateral. Kimura & Sekiguchi stated that the result in the problem is equivalent to each of the following:

(I) The sum $AD + BE + CF$ is not larger than the circumference of the hexagon $AFBDCE$.

(II) For all positive real x, y, and z, one has

$$\sqrt{\frac{y+z}{x}} + \sqrt{\frac{z+x}{y}} + \sqrt{\frac{x+y}{z}} \geqslant 2\left(\sqrt{\frac{x}{y+z}} + \sqrt{\frac{y}{z+x}} + \sqrt{\frac{z}{x+y}}\right).$$

ELEMENTARY PROBLEMS

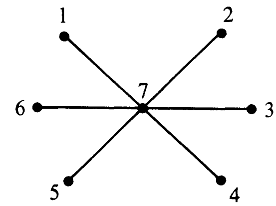
Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by January 31, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2896. *Proposed by Stephen M. Gagola, Jr., Texas A & M University.*

Let m be a positive integer. Prove that $\Sigma\phi(abk)\phi(k) = m^2$, where the sum extends over all triples of positive integers (a, b, k) satisfying $abk \mid m$ and g.c.d. $(a, b) = 1$.

E 2897. *Proposed by David Singmaster, Polytechnic of the South Bank, London, UK.*

Consider a planar graph G and a path P in it. If the path visits a vertex two or more times, we say there is a crossing at the vertex if the path would cross itself when the vertex is viewed as a road intersection; e.g. in the figure on page 538, a path containing $\ldots, 1, 7, 4, \ldots, 2, 7, 5, \ldots$ has a crossing, while a path containing $\ldots, 1, 7, 2, \ldots, 3, 7, 6, \ldots, 4, 7, 5, \ldots$ has no crossing. Prove or disprove: A connected planar graph with every vertex of even degree has an Eulerian circuit with no crossings.



E 2898. *Proposed by U. Abel, Heidelberg KFZ, West Germany.*

A rosette is a directed graph G: $(V \equiv \{x_0, \dots, x_n\}, E)$ with nonzero arc-weights, consisting of cycles which all pass through the central point x_0.

Let A be the matrix corresponding to G, i.e., $a_{ij} =$ weight on (x_i, x_j), and let m_i and n_i denote the number of cycles in G of length $> i$ and $= i$, respectively. Assume that the maximal cycle length l is greater than 1.

Prove the following statements:

(a) $\dim\ker A^k = \sum_{i=1}^{k} (m_j - 1), \qquad k < l;$

(b) $\dim\ker A^l = \dim\ker A^{l-1} = 1 - l + \sum_{k=2}^{l} n_k(k-1).$

E 2899. *Proposed by Gérard Letac, Université Paul-Sabatier, Toulouse, France.*

Let n be a positive integer. Consider a set F of distinct integers such that every element of F is the sum of two elements of F and the sum of k elements of F is never 0 for $k = 1, 2, \dots, n$. Prove that F must have at least $2n + 2$ elements. (*) Find all F with $|F| = 2n + 2$.

E 2900. *Proposed by Edward Neuman, University of Wrocław, Poland.*

The kth divided difference of a function f at the points $t_0, t_1, \dots, t_k$ is denoted by $[t_0, t_1, \dots, t_k]f$ and is defined inductively by

$$[t_0]f = f(t_0), [t_0, t_1, \dots, t_{k+1}]f = ([t_1, t_2, \dots, t_{k+1}]f - [t_0, t_1, \dots, t_k]f)/(t_{k+1} - t_0).$$

Show that, for positive integers k, l,

$$[t_0, t_1, \dots, t_k]x^{k+l} = \sum t_{j_1} t_{j_2} \cdots t_{j_l},$$

the sum being extended over $0 \leqslant j_1 \leqslant j_2 \leqslant \cdots \leqslant j_l \leqslant k$.

E 2901. *Proposed by S. C. Locke and A. Mandel, University of Waterloo.*

Let $f(n) = \gcd\ \{k^n - k \mid k = 2, 3, 4, \dots\}$ for $n \geqslant 2$. Evaluate $f(n)$. In particular, show that $f(2n) = 2$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Sum-Distinct Sets

E 2526 [1975, 300; 1976, 484]. *Proposed by Paul Smith, University of Victoria.*

Call a set $\{a_1, \dots, a_n\}$ of positive integers *sum-distinct* if the 2^n possible sums $\Sigma \varepsilon_i a_i$ (with $\varepsilon_i = 0$ or 1) are all distinct.

Obviously for any n, the set $\{1,2,4,\ldots,2^{n-1}\}$ is an n-element sum-distinct set. Do n-element sum-distinct sets exist with $a_i < 2^{n-1}$ for every i? (For example, $\{3,5,6,7\}$ is a 4-element sum-distinct set with this property.)

Cf. P. Erdős, Problem 220, Canad. Math. Bull., 16 (1973) p. 463.

Comment by Richard K. Guy, The University of Calgary. Solutions and comments appear in [1976, 484] but they don't give the best results. The problem to determine $f(n)$, the least integer which can be the largest element of an n-element sum-distinct set, is an old problem of Erdős. The best-known lower bound is due to Erdős and Leo Moser. The Conway–Guy sequence $u_0 = 0$, $u_1 = 1$, $u_{n+1} = 2u_n - u_{n-r}$ $(n \geqslant 1)$ where r is the nearest integer to $\sqrt{2n}$ generates, at least for small n, the sum-distinct set $\{a_i\}$ where $a_i = u_n - u_i$, $0 \leqslant i \leqslant n-1$, so $f(n) \leqslant u_n$ is a better upper bound than those given. Note that $\lim_{n\to\infty} u_n/2^{n-1} < 0.4702506$.

References

P. Erdős, *Problems and results in additive number theory*, Colloque sur la Théorie des Nombres, Bruxelles, 1955, Liège and Paris, 1956, 127–137, esp. p. 137 with L. Moser.

J. H. Conway and R. K. Guy, *Solution of a problem of P. Erdős*, Colloq. Math., 20 (1969) 307.

An Improper Integral

E 2822 [1980, 220]. *Proposed by R. S. Lehman, University of California, Berkeley.*

Evaluate the limits

(a) $$\lim_{n\to\infty} \int_0^{(2n+1)\pi/2} (\sin t)\ln(1/t)\,dt,$$

(b) $$\lim_{n\to\infty} \int_0^{n\pi} (\cos t)\ln(1/t)\,dt,$$

where n takes on integer values and the integrals are improper Riemann integrals.

Solution by the solvers listed below. Integration by parts gives

(a) $$\int_0^{(n+\frac{1}{2})\pi} (\sin t)\ln(1/t)\,dt = -\ln(n+\tfrac{1}{2})\pi + \int_0^{(n+\frac{1}{2})\pi} \frac{1-\cos t}{t}\,dt$$
$$= \gamma + \int_{(n+\frac{1}{2})\pi}^{\infty} \frac{\cos t}{t}\,dt,$$

since

$$\int_0^x \frac{1-\cos t}{t}\,dt = \gamma + \ln x + \int_x^{\infty} \frac{\cos t}{t}\,dt, \qquad \gamma = \text{Euler's constant}.$$

(b) $$\int_0^{n\pi} (\cos t)\ln(1/t)\,dt = \int_0^{n\pi} \frac{\sin t}{t}\,dt = \frac{\pi}{2} - \int_{n\pi}^{\infty} \frac{\sin t}{t}\,dt.$$

Both infinite integrals go to 0 as n goes to infinity. Thus the answer to (a) is γ and to (b) is $\pi/2$.

Solvers used known properties of the sine- and cosine-integrals, and referred, for example, to I. S. Gradšteĭn and I. M. Ryžik, *Table of Integrals, Series and Products*, Academic Press, 1965, or Abramowitz and Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1965.

The solvers were Ed Adams, W. A. Al-Salam & A. Meir, Miroslav Ašić (Yugoslavia), Theodore S. Bolis, Paul Bracken (Canada), Robert C. Carson, Henry E. Fettis, Franz-Josef Fischer (Germany), John A. Gillespie, David Gootkind, Kenneth Klinger, O. P. Lossers (Netherlands), Helmut Prodinger (Austria), Otto G. Ruehr, Abraham Smuckler (Israel), Milan Šolc (Czechoslovakia), P. D. Vestergaard & S. Berntsen (Denmark), and Paul Vos.

Concave Functions Are Triangle-Preserving

E 2823 [1980, 220]. *Proposed by Herbert Carus, Lanham, Md.*

In Honsberger's *Mathematical Morsels* (see also E 1366 [1959, 432; 1960, 82]) it is shown that the function $f(x) = x^p$, $p = 1/2, 1/3, \dots$, is triangle-preserving, i.e., if x_1, x_2, x_3 are the lengths of the sides of a triangle, then $f(x_1)$, $f(x_2)$, $f(x_3)$ are also the lengths of the sides of a triangle. Show that any (positive) increasing function $f(x)$ for which $f''(x) < 0$ is triangle-preserving.

Solution by O. P. Lossers, Department of Mathematics, Eindhoven University of Technology, Eindhoven, the Netherlands. Since f is positive, nondecreasing, and concave, one has

$$f(y_1 + h) - f(y_1) \geqslant f(y_2 + h) - f(y_2) \qquad (y_1 \leqslant y_2, h \geqslant 0) \tag{1}$$

Let $x_1 \leqslant x_2 \leqslant x_3$. Without loss of generality, $x_1 + x_2 > x_3$; so by setting $h = x_3 - x_2$ one has that $0 \leqslant h \leqslant x_1$. If one now takes $y_1 = x_1 - h$ and $y_2 = x_2$, one obtains

$$f(x_1) \leqslant f(x_2) \leqslant f(x_3)$$

and from (1)

$$f(x_1) + f(x_2) \geqslant f(x_3) + f(x_1 - h) > f(x_3).$$

Remark. Robert A. Melter, Southampton College, located the proposition in *Distance Geometry* by L. M. Blumenthal, Clarendon Press, 1953, pg. 130.

Also solved by Miroslav Ašić (Yugoslavia), Richard Beigel, Robert Breusch, F. S. Cater, Chico Problem Group, Jordi Dou (Spain), Michael W. Ecker, Milton P. Eisner, Lorraine L. Foster, Marshall Fraser, Noel Glick, Doug Hensley, V. Hernandez (Spain), G. A. Heuer, Rodney T. Hood, Brian R. Hunt, Bruce King, L. Kuipers (Switzerland), Joel Levy, James D. McCall, Robert A. Melter, Mark D. Meyerson, James Morrow, Santa Clara Problem Solving Ring, Milan Šolc (Czechoslovakia), Lawrence Somer, David Van Leeuwen, A. Venetoulias & A. Matsoukas (Greece), P. Vos, Daniel Weisser, and David M. Wells.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by January 31, 1982. The solver's full post-office address should be on each sheet.

5980 [1974, 670] **Corrected.** *Proposed by R. M. Cohn, Rutgers University*

Let $M(k)$ be the least upper bound (over distributions) for the probability that a sample of size k from a probability distribution on the positive integers, when ordered, is an arithmetic progression of k distinct integers. Does $M(k)$ exceed $k!/k^k$, which is the value of the probability when the distribution assigns probability $1/k$ to the first k integers? Show that $\lim_{k\to\infty} M(k) = 0$.

6350. *Proposed by James Wiegold, University College, Cardiff, Wales.*

Prove that the alternating group $A_n (n > 4)$ can be generated by three involutions. Can two of them be conjugate? (*) Can one of them be preassigned?

6351. *Proposed by James W. Walker, MIT.*

Let F be a closed subset of the unit square $I^2 = [0,1] \times [0,1]$ such that F contains the closed horizontal edges $[0,1] \times \{0\}$ and $[0,1] \times \{1\}$, but does not intersect the open vertical edges $\{0\} \times (0,1)$ and $\{1\} \times (0,1)$. Show that either both horizontal edges belong to the same connected component of F, or else both vertical edges belong to the same component of $I^2 \setminus F$.

6352. *Proposed by J. C. Lagarias, Bell Laboratories.*

Let $\|\cdot\|$ denote the Euclidean norm on R^2. Given four points x_i in R^2 that satisfy

(i) x_i is in the closure of the ith quadrant of R^2,
(ii) $\|x_i\| \leqslant 1$ for all i,
(iii) $\|x_i + x_{i+1}\| > 1$ for $1 \leqslant i \leqslant 4$, where $x_5 = x_1$.
Show that $\|x_1 + x_2 + x_3 + x_4\| < 1$. Is it true that if (iii) is replaced by $\|x_i + x_{i+1}\| \geqslant 1$ then $\|x_1 + x_2 + x_3 + x_4\| \leqslant 1$?

6353. *Proposed by the editors.*

"Everyone" knows that the symmetric group S_n can be embedded in A_{n+2}. Can S_n be embedded in A_{n+1} (alternating group)?

6354. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that there exists a constant $K > 0$ such that

$$\|(tI - A)^{-1}\| \leqslant \frac{K}{t - r} \quad \text{for all} \quad t > r.$$

Here I denotes the identity matrix and $\| \ \|$ is the operator norm induced by the Euclidean norm $x \to (x^T x)^{1/2}$.

*Find the best value of the constant K.

6355. *Proposed by M. J. Pelling, Watford, England.*

For what pairs m, n of positive integers do there exist two rational irreducible polynomials $f(x), g(x)$ of the same degree $m + n$ such that if u is a root of $f(x)$ then, over $Q(u)$, $g(x)$ factors into the product of an irreducible polynomial of degree m and an irreducible polynomial of degree n? (For example 3, 4 is a solution but 2, 3 is not.)

SOLUTIONS OF ADVANCED PROBLEMS

Translation Invariance Hamel Basis

6278 [1979, 709]. *Proposed by Stanley Wagon, Smith College, Northampton, Mass.*

Let X be the real vector space consisting of all bounded real-valued functions on the reals with bounded support. Is there a basis, B, for X which is closed under translation, i.e., if f is in B and t is real, then f_t is in B where $f_t(x) = f(x + t)$? (An affirmative answer would provide a new proof of Banach's theorem that a non-trivial, finitely additive, translation invariant measure defined for all bounded sets of reals exists.)

Solution by Leroy F. Meyers, The Ohio State University. The notation is adjusted so that f_t is the translate of f *to the right* by t, i.e., $f_t(x) = f(x - t)$ for all f, x, and t. Superscripts are markers, not exponents. Note that for no nonzero f and t can we have $f_t = f \in X$.

Suppose that there is a translation-invariant Hamel basis for X. Let f^0 be the characteristic function of the half-open unit interval $[0, 1)$, and let H^0 be the set of basis elements occurring with nonzero coefficient in the expansion of f^0. Since H^0 is a finite set, we may fix an integer n such that no pair $\{h, h_{p+(1/n)}\}$ with integral p is a subset of H^0. Let g^0 be the characteristic function of $[0, 1/n)$, and let K^0 and K_t^0 be the sets of basis elements occurring with nonzero coefficient in the expansions of g^0 and g_t^0, respectively. Obviously, $n > 1$ and $\Sigma_{j=0}^{n-1} g_{j/n}^0 = f^0$.

Suppose that $k \geqslant 0$ and that f^k, g^k, H^k, K^k, and K_t^k have been defined so that $f^k = \Sigma_{j=0}^{n-1} g_{j/n}^k$, H^k, K^k, and K_t^k are the sets of basis elements occurring with nonzero coefficient in the expansions of f^k, g^k, and g_t^k, respectively, and no pair $\{h, h_{p+(1/n)}\}$ with integral p is a subset of H^k. If $g^k = 0$, then set $m = k$ and proceed to the next paragraph. Otherwise, let h^k be a minimal element of K^k in the sense that no other element of K^k has a support with smaller infimum. Then $h^k \in H^k$, since $h^k \notin K_{j/n}^k$ for $j > 0$. Let $a_k h^k$ be the term using h^k in the expansion of g^k. Then

$a_k \neq 0$ and $a_k h^k_{j/n}$ is the term using $h^k_{j/n}$ in the expansion of $g^k_{j/n}$ for $0 \leqslant j \leqslant n-1$. Now $h^k_{1/n} \notin H^k$, and also $h^k_{1/n} \notin K^k_{j/n}$ for $j \geqslant 2$. Hence the "cancelling" term $-a_k h^k_{1/n}$ must occur in the expansion of g^k. We now set $g^{k+1} = g^k - (a_k h^k - a_k h^k_{1/n})$ and $f^{k+1} = \Sigma_{j=0}^{n-1} g^{k+1}_{j/n} = \Sigma_{j=0}^{n-1} g^k_{j/n} - a_k \Sigma_{j=0}^{n-1} (h^k_{j/n} - h^k_{(j+1)/n}) = f^k - a_k(h^k - h^k_1)$, and let H^{k+1}, K^{k+1}, and K^{k+1}_t be the sets of basis elements occurring with nonzero coefficient in the expansions of f^{k+1}, g^{k+1}, and g^{k+1}_t, respectively. Now $K^{k+1} = K^k \setminus \{h^k, h^k_{1/n}\}$ and H^{k+1} differs from H^k only in the possible insertion or deletion of h^k or h^k_1. Hence K^{k+1} is a proper subset of K^k, and no pair $\{h, h_{p+(1/n)}\}$ with integral p is a subset of H^{k+1}. Thus the induction step is complete.

The definition process must terminate, since K^0 is a finite set. Now $f^m = \Sigma_{j=0}^{n-1} g^m_{j/n} = 0$. Hence $f^0 = \Sigma_{k=0}^{m-1} (f^k - f^{k+1}) = \Sigma_{k=0}^{m-1} a_k(h^k - h^k_1) = F - F_1$, $F = \Sigma_{k=0}^{m-1} a_k h^k \in X$. Now $F(x) - F(x-1) = f^0(x) = 1$ if $[x] = 0$ and $F(x) - F(x-1) = f^0(x) = 0$ otherwise. Hence $F(x) = F(\langle x \rangle)$ if $x \geqslant 0$ and $F(x) = F(\langle x \rangle) - 1$ if $x < 0$, where $\langle x \rangle = x - [x]$. Since $F(\langle x \rangle)$ and $F(\langle x \rangle) - 1$ cannot both be 0 for the same x, F cannot have bounded support, and so $F \in X$. Hence no translation-invariant Hamel basis exists.

A Condition on Entire Functions

6279 [1979, 793]. *Proposed by Lee A. Rubel, University of Illinois, Urbana-Champaign.*

Let $f(z)$ be an entire function such that the maximum modulus over every closed line segment L is achieved at one of the endpoints a and b of L; that is,

$$\max\{|f(z)| : z \in L\} = \max\{|f(a)|, |f(b)|\}.$$

Prove that $f(z)$ has either the form $A(z-B)^n$ or the form $A \exp Bz$, where A and B are constants and n is a nonnegative integer.

Composite of solutions by P. R. Chernoff and J. Essick, University of California, Berkeley, and by Jan Malý and Zdeněk Vlášek, Charles University, Prague. If f has more than one zero, then $|f|$ vanishes on the segment joining two of its zeros; so f is identically zero. Assume henceforth that f is not identically zero.

Case 1. f has one zero. Without loss of generality we may suppose $f(0) = 0$. If f is a polynomial, it is necessarily of the form cz^n. But if f is not a polynomial, then by the Casorati-Weierstrass theorem there is a sequence of points p_n with $|p_n| \to \infty$ and $|f(p_n)| = \varepsilon_n \to 0$. By passing to a subsequence, we may suppose that the unit complex numbers $\omega_n = p_n/|p_n|$ converge to a limit ω. Then $|f(z)| \leqslant \varepsilon_n$ on the line segment joining 0 to p_n, so that $|f(z)|$ must be identically 0 on the ray through 0 along direction ω. This is a contradiction.

Case 2. f has no zeros. Then $f = e^g$ for an entire function g. The real part $u(z)$ of $g(z)$, when restricted to any line L, can have no local maxima. Consequently, if the directional derivative $D_\omega u(a)$ along the direction ω at the point a should be positive, it follows that $u(a + t\omega)$ must be nondecreasing for $t \geqslant 0$, and thus $u(z) \geqslant u(a)$ for all z on the ray emanating from a along direction ω. Note that $D_\omega u(a) = \mathrm{Re}(\omega g'(a))$. Hence $\mathrm{Re}(\omega g'(a)) \leqslant 0$, and thus the range of g' lies in a half-plane. Therefore g' is a constant A, $g(z) = Az + B$, and $f(z) = \exp(Az + B)$.

Also solved by Dieter Schmersau and the proposer.

MISCELLANEA

59. "I owe it to mathematics alone that I know how to control my thoughts; this study tamed and cooled down my fantasy which formerly dominated me without restraint; by subjecting it to reason it doubled its power."—Antoine Reicha (Antonín Rejcha), quoted in the liner notes accompanying Crossroads phonograph record 22 16 0110. (Reicha lived 1770–1836.) (Suggested by John W. Dawson.)

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

Precalculus, T(13: 1). College Algebra. Jimmie Gilbert, James Spencer, Linda Gilbert. P-H, 1981, xv + 432 pp, $16.95. [ISBN: 0-13-141804-1] Excellent combination of readability and thoroughness. Provides motivation and intuitive understanding along with careful statements of principles. More than enough material for a one semester course. Answers to odd-numbered exercises. Solutions provided for all problems in end-of-chapter practice tests. MB

Precalculus. The Power of Mathematics: Applications to Management and the Social Sciences, Second Edition. Kenneth L. Whipkey, Mary Nell Whipkey, George W. Conway, Jr. Wiley, 1981, x + 622 pp, $21.95. [ISBN: 0-471-07709-7] One more offering of what might constitute the mathematics suitable for students in management and social sciences: some linear algebra and linear programming, some statistics and probability, a chapter on business mathematics, and two chapters on calculus. (First Edition, TR, June-July 1979.) AWR

Precalculus, T*(13: 1). College Algebra, Third Edition. Margaret L. Lial, Charles D. Miller. Scott F, 1981, xvi + 429 pp, $16.95. [ISBN: 0-673-15407-6] Changes from second edition (First Edition, TR, October 1973; Second Edition, TR, August-September 1978) include: extensive revision of chapters on graphing, greater range of difficulty in exercises, calculator exercises. Text includes some topics not ordinarily covered, e.g., matrices, sequences, counting methods. JG

Precalculus, T(13, 1). Trigonometry, Second Edition. Margaret L. Lial, Charles D. Miller. Scott F, 1981, xiv + 322 pp, $16.95. [ISBN: 0-673-15432-7] Second edition of a nicely produced standard text for trigonometry (First Edition, TR, November 1977). Effort is made to incorporate the calculator, though this is made difficult by the need to know about inverse trigonometric functions to use a calculator for material that traditionally (and in this text) comes early. AWR

Precalculus, T(13: 1). College Algebra and Trigonometry. Robert Ellis, Denny Gulick. HarBrace J, 1981, x + 661 pp, $18.95. [ISBN: 0-15-507907-7] More than enough material for a semester course. Combines readability and clarity with thoroughness and accuracy. Much use of illustrative examples. Ample problem sets, including some calculator exercises. Answers to odd-numbered problems. Answer manual available for even-numbered problems. MB

Precalculus, T(13: 1), L. Essentials of Trigonometry, Third Edition. Irving Drooyan, Walter Hadel, Charles C. Carico. Macmillan, 1981, viii + 328 pp, $16.95. [ISBN: 0-02-330270-4] Some textual rewriting and re-organization to utilize calculators, but basically the same book as earlier editions (First Edition, TR, October 1971; Second Edition, TR, January 1978). Straightforward presentation for a one-term trigonometry course, starting with right triangles and ending with complex numbers and polar coordinates. Includes Laws of Sines and Cosines, Hero's formula, vectors, nth roots. Exercises (with answers), index, and appendix on Algebra and Geometry Review. JS

Foundations, P. Formal Language Theory: Perspectives and Open Problems. Ed: Ronald V. Book. Acad Pr, 1980, xiii + 454 pp, $25. [ISBN: 0-12-115350-9] Texts of 13 of the 16 lectures presented at the Symposium on Formal Language Theory that was held in Santa Barbara, California, December 10-14, 1979. RJA

Foundations, P. Initial Segments of Degrees Below 0′. Richard L. Epstein. Memoirs No. 241. AMS, 1981, vi + 102 pp, $6 (P). [ISBN: 0-8218-2241-1]

Number Theory, T*(17: 1), S, P*, L. Analytic Number Theory, An Introduction. Richard Bellman. Benjamin/Cummings, 1980, xvi + 195 pp, $19.50. [ISBN: 0-8053-0360-X] An introduction to analytic number theory which focuses on the mean value of certain elementary arithmetic functions. The text is sparse, but there are lots of challenging problems and excellent bibliographical references. CEC

Linear Algebra, T(13-14: 1). Elementary Linear Algebra and Its Applications. James W. Daniel. P-H, 1981, xii + 338 pp, $18.95. [ISBN: 0-13-258293-7] A text which has grown out of the author's collaboration with Ben Noble. Good sophomore level treatment of applications--aimed directly at the student. LLK

Linear Algebra, T(14: 1). Basic Linear Algebra with Applications. Garfield C. Schmidt. Krieger, 1980, xiii + 521 pp, $34.50. [ISBN: 0-89874-000-2] Nice treatment of topics and their applications but one would need to choose carefully which to fit into a semester course. For example: quadratic forms are used to determine critical points of functions of several variables. Outrageous price! LLK

Algebra, S(18), P. Algebraic Structures of Symmetric Domains. Ichiro Satake. Pub. of Math. Soc. of Japan, No. 14. Princeton U Pr, 1980, xvi + 321 pp, $39.50. An attempt to give a unified treatment of structures related to symmetric domains; including Lie groups, Jordan algebras, and Siegel domains. Though intended for non-specialists and including introductory chapters the presentation is fast-paced and sophisticated; a sound foundation in Lie groups is essential. Approach is by way of morphisms, leading to results of Wolf, Koranyi, Kuga, and the author. Exercises (difficult), appendices, indexes, and extensive bibliography. JS

Algebra, T*(17-18: 1, 2), P, L. Category Theory, An Introduction, Second Edition. Horst Herrlich, George E. Strecker. Heldermann Verlag, 1979, xiii + 400 pp, (P). [ISBN: 3-88538-001-3] This Second Edition (First Edition, TR, March 1974) refines and corrects the presentation of introductory category theory given in the first edition. The already very extensive bibliographies (Books, Proceedings, and Papers) are updated. These, together with many exercises and a nearly 20-page index, make this book valuable both as a text and as a reference work for the non-specialist. JAS

Algebra, P. Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebras. W.G. McKay, J. Patera. Lect. Notes in Pure and Appl. Math., V. 69. Dekker, 1981, v + 317 pp, $39.75 (P). [ISBN: 0-8247-1227-7] Representations of complex simple Lie algebras are described in terms of representations of maximal semi-simple subalgebras by means of tables (branching rules). There are 24 pages of text and references followed by nearly 300 pages of tables. Rank is limited to 8 or less; dimensions to 5,000 for the classical algebras; 10,000 for the exceptional ones. JS

Calculus, T(13-14: 3), L. Calculus and Analytic Geometry. Philip Gillett. Heath, 1981, xvii + 906 pp, $28.95. [ISBN: 0-669-00641-6] Solutions Guide, Volume 1, 247 pp, (P) [ISBN: 0-669-00642-4]; Solutions Guide, Volume 2, 238 pp, (P) [0-669-03212-3]. Written to provide motivation and intuitive understanding along with clarity and precision. Students should find it readable and interesting. Ample problem sets, with answers to odd-numbered exercises, and to all questions in end-of-chapter true-false quizzes. Thorough index. MB

Calculus, S(13). How To Enjoy Calculus, Revised Edition. Eli S. Pine. ARCO Pub, 1980, 132 pp, $4.95 (P); $7.95. [ISBN: 0-668-04951-1; 0-668-04949-9] Probably useful for a large number of students as a simplified calculus. (First Edition, TR, March 1978.) LLK

Calculus, S*(13-14). Mathematical Handbook, Higher Mathematics, Third Printing. M. Vygodsky. Trans: George Yankovsky. MIR Pub, 1978, 872 pp, $12. Continuation of the Handbook of Elementary Mathematics by the same author. Reference work covering plane and solid analytical geometry, calculus of one and several variables, and ordinary differential equations, with eighty pages on "some remarkable curves." Not intended to replace a textbook. Numerous examples but no problems. Some proofs, with emphasis on content. Historical notes. Excellent name and subject indexes. Brief 4-place tables of the usual functions. (Second Printing, TR, August-September 1976.) JK

Calculus, T(13-14: 3). The Calculus with Analytic Geometry, Fourth Edition. Louis Leithold. Har-Row, 1981, xvi + 1200 pp, $28.95. [ISBN: 0-06-043935-1] Retains thoroughness and readability of earlier editions, but many sections rewritten. Keeps definition of functions as sets of ordered pairs--now applies this in defining function inverses. 2000 new exercises, some involving metric units. Careful treatment of series. Uses green for emphasis both in text and drawings. Ample material for three semesters. (TR, First Edition, June-July 1969; Second Edition, June-July 1972 and January 1973; Third Edition, June-July 1976.) MB

Complex Analysis, T(17), S. Conformal Mapping on Riemann Surfaces. Harvey Cohn. Dover, 1980, xiv + 325 pp, $6 (P). [ISBN: 0-486-64025-6] A corrected reprint of the 1967 edition (TR, February 1968). An extended review of that edition appears in the December 1968 Monthly. TAV

Differential Equations, T(14-15: 1, 2), L. Elementary Differential Equations, Sixth Edition. Earl D. Rainville, Phillip E. Bedient. Macmillan, 1981, xiv + 529 pp, $21.95. [ISBN: 0-02-397770-1] A Short Course in Differential Equations, Sixth Edition. Macmillan, 1981, xi + 335 pp, $18.95. [ISBN: 0-02-397760-4] Main change from previous edition (TR, Fourth Edition, June 1969; Fifth Edition, TR, December 1974) is the increase in variety and number of applications, and earlier consideration of some applications. Retains the readability of earlier editions. Contains considerably more than enough material for a semester course. A Short Course consists exactly of the first sixteen chapters of Elementary Differential Equations, omitting larger book's consideration of power series and numerical methods, partial differential equations, Fourier series, and boundary value problems. MB

Differential Equations, S(15-16), L. Special Functions of Mathematical Physics and Chemistry, Third Edition. Ian N. Sneddon. Longman, 1980, ix + 182 pp, $14.50 (P). [ISBN: 0-582-44396-2] More thorough treatment of hypergeometric, Legendre, Bessel, Hermite, and Laguerre functions than is found in differential equations texts. Good problem sets. This edition brings in contour integration as an additional approach to solving differential equations. MB

Differential Equations, T*(17: 1), P.** Theory and Applications of Hopf Bifurcation. B.D. Hassard, N.D. Kazarinoff, Y.-H. Wan. London Math. Soc. Lect. Note Ser., No. 41. Cambridge U Pr, 1981, 311 pp, $35 (P). [ISBN: 0-521-23158-2] Theory of the not uncommon phenomenon of the birth of a family of oscillations as a controlling parameter is varied. Algorithms for computing the bifurcating periodic solutions, their periods, and stability. Computer programs provided on microfiche. Numerous worked examples such as Watt's steam engine governor and the Hodgkin-Huxley model nerve

conduction equations. Unstuffy, clear exposition on uncrowded reproduced typewritten pages. Over 100 selected references to books and journal articles. Brief index. Exercises for each of the five chapters. Recommended for engineers and scientists as well as mathematicians. JK

Differential Equations, T?(18: 2), P. Nonlinear Differential Equations. Svatopluk Fucik, Alois Kufner. Stud. in Appl. Mech., No. 2. Elsevier Sci Pub, 1980, 359 pp, $83. [ISBN: 0-444-99771-7] Written to appeal to engineers as well as mathematicians. Exposition of principal ideas and methods, not an exhaustive survey. Needed mathematics explained at point of application to show interconnections among disciplines, especially functional analysis. Most proofs not given in detail. Specific equations of mathematical physics are omitted. Illustrative examples, which require reader's active participation, serve as exercises. Over 90 references. Good index. JK

Numerical Analysis, P. Optimal Quadrature Formulas. Meishe Levin, Jury Girshovich. B.G. Teubner, 1979, 124 pp, (P). Basic results concerning the construction of one- and two-dimensional optimal quadrature formulas (in explicit form) for sets of functions with given degree of smoothness. Methods use monosplines and their properties. Chapter on asymptotically optimal quadrature formulas. Bibliography with 66 entries, many in Russian. JK

Numerical Analysis, P. Colloquium Numerical Solution of Partial Differential Equations. Ed: J.G. Verwer. MC Syllabus, No. 44. Math Centrum, 1980, iv + 194 pp, Dfl. 24 (P). [ISBN: 90-6196-205-6] Reports of current research in the numerical solution of partial differential equations. Presented at six one-day meetings one month apart held, in turn, in Delft, Nijmegen and Amsterdam. Topics include multi-grid methods, incomplete factorization and pre-conditioned iterative methods, shallow-water equations, Navier-Stokes equations, free boundary problems, numerical solution of parabolic equations and splitting methods for time-dependent problems. JK

Numerical Analysis, P. Endliche Lagstrukturen. Roland Fahrion. Physica-Verlag, 1980, 229 pp, (P). [ISBN: 3-7908-0226-3] An account for specialists of the theory of spline lags. JD-B

Optimization, P, L. Simplicial Fixed Point Algorithms. G. Van der Laan. Math. Centre Tracts, No. 129. Math Centrum, 1980, iv + 172 pp, Dfl. 21 (P). Survey of known algorithms and presentation of two new ones—a restart algorithm which generates a path of simplices of variable dimension, and one which automatically refines grid size. Discussion of convergence conditions. JRG

Analysis, P. Padé-Type Approximation and General Orthogonal Polynomials. Claude Brezinski. Int. Ser. Num. Math., V. 50. Birkhäuser, 1980, 250 pp, $34. [ISBN: 3-7643-1100-2] Padé approximants provide rational approximations to functions which are formally defined by power series. By introducing Padé-type approximants and studying these on the basis of general orthogonal polynomials, the author achieves a unified algebraic theory of Padé approximants. TRS

Probability, P. Censoring and Stochastic Integrals. R.D. Gill. Math. Centre Tracts, No. 124. Math Centrum, 1980, v + 178 pp, Dfl. 22 (P). [ISBN: 90-6196-197-1] Censoring in the title refers to the (probabilistic) elimination of certain data from a sample for reasons either under or outside the experimenter's control. Stochastic processes experiencing censoring require special models and this monograph pulls together many of the known techniques. Contains an extensive bibliography. TAV

Probability, T?(17), S, P. An Introduction to Vector Stochastic Processes. Kenneth S. Miller. Krieger, 1980, vii + 214 pp, $19.50. [ISBN: 0-88275-855-1] Assuming a background in matrix theory, differential equations, and probability, the author presents a self-contained treatment of several central ideas in vector stochastic processes. An accessible source book for anyone wishing to learn the theory. No exercises, no applications. TAV

Probability, T(17: 2), P. An Introduction to Stochastic Processes and Their Applications. Chin Long Chiang. Krieger, 1980, xxi + 517 pp, $36.50. [ISBN: 0-88275-200-6] Intended as a text in stochastic processes or probability theory. Take seriously the Preface note that "A basic knowledge in probability and statistics is required for a profitable reading of the text." The first eight chapters are revised from a 1968 book by the author; nine new chapters are included. AWR

Probability, P, L*. Handbook of Applicable Mathematics, V. II: Probability. Emlyn Lloyd. Wiley, 1980, xix + 450 pp, $85. [ISBN: 0-471-27821-1] Second of six core volumes in this new series (for description of series see TR, December 1980, of Volume I: Algebra). Covers basic concepts of probability theory and stochastic processes. RSK

Statistics, T(13-15: 1, 2). Statistics for Business and Economics. William R. Heitzman, Frederick W. Mueller. Allyn, 1980, xv + 670 pp, $17.95. [ISBN: 0-205-06753-0] The usual topics plus additional material on regression analysis, decision theory, index numbers, and time series. Presupposes only school mathematics. FLW

Statistics, P. Mathematical Statistics. D.R. Brillinger, et al. Selecta Statistica Canadiana, V. 5. U Pr of Canada, 1979, 150 pp, (P). An annual volume containing seven research papers. LAS

Statistics, T(15-17), P. Basic Statistics for Social Research, Second Edition. Dean J. Champion. Macmillan Pub, 1981, xii + 452 pp, $19.95. [ISBN: 0-02-320600-4] Although this book is written for a noncalculus audience, the level of writing, the relative sparseness of exercises, and the discussions of appropriateness of methods for different situations all mark the book as more sophisticated

than the standard introductory text. It can function, as claimed, as a source book for those primarily interested and trained in social science. AWR

Statistics, T(13-14: 1). Introductory Statistics for the Behavioral Sciences, Fourth Edition. Robert K. Young, Donald J. Veldman. HR&W, 1981, x + 687 pp, $18.95. [ISBN: 0-03-043051-8] In this new edition of this programmed text for self-paced study (Second Edition, TR, February 1974; Third Edition, TR, March 1978), the order has been changed a bit and there is new material on regression and non-parametric tests. FLW

Computer Programming, T(13-15: 1), L. Programming Standard PASCAL. R.C. Holt, J.N.P. Hume. Reston, 1980, x + 381 pp, $11.95 (P). [ISBN: 0-8359-5690-3] Thorough introduction to programming using Pascal. Divides Pascal's features into successive "Pascal Subsets." Also contains chapters on numerical methods, other programming languages, assembly and machine languages, and compilers. Quite readable. MB

Computer Programming, S(15-18), P. The New UCI LISP Manual. Ed: James R. Meehan. Lawrence Earlbaum Assoc, 1979, xix + 366 pp, $14.95 (P). [ISBN: 0-89859-012-4] Present text is a unification and reworking of the Stanford LISP Manual (SAILON 28.7), the UCI LISP Manual (Tech Report 21, October 1973), and the Rutgers/UCI LISP Manual (April 1978). Index. RJA

Computer Science, T(15-18: 1, 2), S, L. Operating Systems. Harold Lorin, Harvey M. Deitel. A-W, 1981, xxi + 378 pp, $19.95 (P). [ISBN: 0-201-14464-6] Text emphasizes the relationship between operating systems and hardware and other systems software; the interface to users; the intersection of operating system design and software technology; the fundamental algorithms of resource management; the latest, nontraditional directions in operating system research and technology. Attempts to convey the present state of seeking new, alternate answers to basic questions concerning what an operating system is and what it should do. Bibliography. Index. RJA

Computer Science, T(13-18: 1, 2), S, L. LISP. Patrick Henry Winston, Berthold Klaus Paul Horn. Addison-Wesley, 1981, xii + 430 pp, $11.95 (P). [ISBN: 0-201-08329-9] Part one introduces the basics of programming in LISP. The second part illustrates the use of LISP in artificial intelligence and related fields. Included in part two are chapters on the Blocks World, symbolic pattern matching and simple theorem proving, interpreting and compiling augmented transition networks, program writing programs and natural language interfaces. Chapter problems and answers. Appendexes. Bibliography. Index. RJA

Computer Science, T*(14-18: 1, 2). Syntax of Programming Languages: Theory and Practice. Roland C. Backhouse. Prentice-Hall, 1979, xv + 301 pp, $25.95. [ISBN: 0-13-879999-7] Provides the bridge between the practice of compiler construction and the theory of computing. Begins with fundamentals on context-free grammars and graph searching. Includes chapters on the definitions of regular languages, LL and LR parsing, error repair and recovery. Chapter exercises. Chapter bibliographic notes. Index. RJA

Computer Science, T(15-17: 1), S, P. Digital Computer Simulation. Fred Maryanski. Hayden Book, 1980, 328 pp, $15.95. [ISBN: 0-8104-5118-2] A well-written introductory text on simulation for students who know higher-level language programming and some calculus. Simulation programming as well as the design and analysis of simulation experiments are dealt with. Both discrete system simulation, using GPSS and Simscript, and continuous simulation, using CSMP and DYNAMO, are treated. With exercises and a chapter on probability-statistics background material. JL

Computer Science, T*(15-18: 1, 2), S, P, L. Artificial Intelligence Programming. Eugene Charniak, Christopher K. Riesbeck, Drew V. McDermott. Lawrence Earlbaum Assoc, 1980, xii + 323 pp, $19.95. [ISBN: 0-89859-004-3] Provides commonly used tools for programming artificial intelligence theories: discrimination nets, agendas, deduction, data dependencies, backtracking, etc. First part is on advanced LISP programming; second covers artificial intelligence programming techniques; third presents a sample artificial intelligence project. Included are LISP implementations of ideas presented in part two. Some exercises. Appendix. Bibliography. Indexes. RJA

Computer Science, T(14-18: 1, 2), S, P, L. 16-Bit Microprocessor Architecture. Terry Dollhoff. Reston Pub, 1979, xv + 471 pp, $24.95. [ISBN: 0-8359-7001-9] Presents the technology that produced the new 16-bit chips and their associated hardware and software. Part one provides historical information on microprocessors and discusses the more popular 8-bit machines. The second part contains a detailed case study of the 16-bit 9900 microprocessor. The last part is devoted to an overview of other 16-bit microprocessors: the 8086, the Z8000, Nova compatible micros, PACE, and the Motorola 68000. Appendexes. Index. RJA

Computer Science, T(14-15: 1, 2), L. Fundamental Structures of Computer Science. William A. Wulf, et al. A-W, 1981, xviii + 621 pp, $16.95. [ISBN: 0-201-08725-1] A text that relates programming techniques with the underlying mathematical concepts (automata, data types, etc.). Major portions of the book are devoted to control structures, data structures, and their interaction. Several chapter-length examples are presented at the end. AO

Computer Science, T(14-15: 1), P. Introduction to Computer Design and Implementation. S. Imtiaz Ahmad, Kwok T. Fung. Computer Sci Pr, 1981, ix + 271 pp, $19.95. [ISBN: 0-914894-11-0] Presents the basic principles of computer design, starting from the design of simple combinational and sequential circuits. Implementation using state-of-the-art technology is also discussed. AO

Computer Science, T*(14-15: 1), P, L*. A Structured Programming Approach to Data. Derek Coleman. Springer-Verlag, 1979, xi + 222 pp, $12.10 (P). [ISBN: 0-387-91138-3] This text presents the theory of structured programming with a particular focus on the design of programs in terms of appropriate abstract data types. The representation of these data structures on real computers is also discussed. Well-chosen exercises. AO

Computer Science, T(16-18: 1), S, P. The Definition of Programming Languages. Comp. Sci. Texts, No. 11. Cambridge U Pr, 1980, xii + 268 pp, $29.95; $12.95 (P). [ISBN: 0-521-22631-7; 0-521-29585-8] This book covers the main developments in programming language definitions over the years. Definitions and methods of definition of the major programming languages are described and evaluated in detail. Very useful as a reference or as a supplementary text. JL

Applications (Data Analysis), P. Pratique de L'Analyse des Données. J.-P. Benzécri, F. Benzécri. Dunod, 1980. T. 1: Analyse des correspondances exposé élémentaire, viii + 424 pp, (P). [ISBN: 2-04-011227-8]; T. 2: Abrégé théorique études de cas modèle, xi + 466 pp, (P). [ISBN: 2-04-011181-6] An introduction to the analysis of data; a thorough presentation of linear algebraic preliminaries (Volume 1), followed by numerous case studies (Volume 2). SG

Applications (Data Processing), T(13: 1), S, L. Introduction to Computer Data Processing, Third Edition. Wilson T. Price. HR&W, 1981, xiv + 577 pp, $18.95. [ISBN: 0-03-056728-9] A beginning text on data processing that provides a non-technical introduction (with exercises) to how a computer works, to programming and to business applications. JL

Applications (Economics), T(13-14: 1, 2), L. Mathematical Applications for Management, Life, and Social Sciences. Ronald J. Harshbarger, James J. Reynolds. Heath, 1981, xiv + 608 pp, $17.95. [ISBN: 0-669-03209-3]; Selected Solutions, 155 pp, $2.95 (P) [ISBN: 0-669-03211-5] A mixture intended for students in the management, life, and social sciences; combining elementary linear algebra, probability, and calculus. Includes simplex method for linear programming, mathematics of finance, some statistics, calculus through partial derivatives and Lagrange multipliers. Develops some concepts, but mostly emphasizes formulas and manipulation. Exercises (with answers); index. Solutions manual available. JS

Applications (Economics), P. Multinomial Probit: The Theory and Its Application to Demand Forecasting. Carlos Daganzo. Acad Pr, 1979, xiv + 222 pp, $30.50. [ISBN: 0-12-201150-3] Multinomial probit (MNP), a general discrete choice model, is a counterpart of multiple regression that can handle categorical data for econometric modelling. Five chapters cover maximum likelihood estimation, calibration, prediction and statistical interpretation; four appendices cover necessary mathematical background. LAS

Applications (Engineering), P. Mathematical Aspects of Marine Traffic. Ed: S.H. Hollingdale. Acad Pr, 1979, xii + 269 pp, $47. [ISBN: 0-12-352450-4] A compilation of 17 papers presented at a conference at London University in 1977. Narrowly focused, as the title indicates, it contains some interesting applications of mathematical modelling and operations research. AWR

Applications (Physics), S(17-18), P. Techniques and Applications of Path Integration. L.S. Schulman. Wiley, 1981, xv + 359 pp, $31.95. [ISBN: 0-471-76450-7] Provides an introduction to the path integral and surveys a number of important applications in diverse areas of physics. AO

Applications (Physics), S(17-18), P. Mathematical Theory of Elastic and Elasto-Plastic Bodies: An Introduction. Jindrich Necas, Ivan Hlavácek. Stud. in Appl. Mech., No. 3. Elsevier Sci Pub, 1981, 342 pp, $73.25. [ISBN: 0-444-99754-7] In addition to presenting the basic theory of elasticity and plasticity, this book also reviews the contemporary state of the theory. Only static problems are considered and primary emphasis is placed on variational techniques. AO

Applications (Software Engineering), P. Software Engineering. Ed: Herbert Freeman, Philip M. Lewis II. Acad Pr, 1980, x + 244 pp, $21. [ISBN: 0-12-267160-0] Proceedings of the Software Engineering Workshop held in Albany, Troy, and Schenectady, New York from May 30-June 1, 1979. Workshop was co-sponsored by the General Electric Research and Development Center, the Rensselaer Polytechnic Institute, and the National Science Foundation. Index. RJA

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

MARCH MEETING OF THE FLORIDA SECTION

Dedicated to the Memory of James D. McKnight

The Fourteenth Annual Spring Meeting of the Florida section of MAA was held on March 13-14, 1981, at Bethune-Cookman College. One-hundred sixty-six persons registered their attendance.

Seven invited addresses were presented as follows: Dr. Alfred B. Willcox, Executive Director of MAA, "Some Bridges to and From Mathematics or There is a Mathematician Loose in the Supermarket;" Professor *Lamberto Cesari* of the University of Michigan, "Calculus of Variations: A Twentieth Century View;" Professor *Richard D. Anderson,* Louisiana State University, "Some Algorithmically Defined Functions;" Professor *Harley Flanders,* Florida Atlantic University, "Some Geometry From My Museum;" Professor *John Mayer,* University of Florida, "The Fixed Point Property and Embedding of Non-separating Plane Continua;" Professor *Christopher Hunter,* Florida State University, "Matthieu's Equation and Its Eigenvalues;" Professor *Steven Shafer* of Carnegie Mellon University, "Geometry in Visual Perception."

In conjunction with the meeting there was a State Articulation Conference. The following talks and panel discussions were presented: "Fourth International Congress on Mathematical Education." Panel: *Don Hill,* Florida A&M University, Moderator; *Don Lichtenberg,* University of South Florida; *Gene Nichols,* Florida State University; *Betty Lichtenberg,* University of South Florida; *Ray Roth,* Rollins College; *Elizabeth Magarian,* Stetson University. "Essential Academic Skills for Florida Community Colleges: *Ernie Ross,* Clearwater Campus of St. Petersburg Junior College, and *Etta Mae Whitten,* Tallahassee Community College. "Articulation Update - Quarters to Semesters," Panel: Members of the Committee for Articulation in Mathematics. Presiding, *Bill Rice,* St. Petersburg Junior College.

A Saturday morning session, sponsored by Pi Mu Epsilon, featured the following talks: *Mark Bateh* Florida Delta Chapter, University of Florida, "The QR Algorithm for Finding Eigenvalues;" *Don Sharpe,* Georgia Epsilon Chapter, Valdosta State College, "Huffman Encoding;" *Michael Walloga,* Florida Epsilon Chapter, University of South Florida, "The Integration of Sec^3 x dx."

The following papers were presented to the section: *Withold Kosmala,* University of Tampa,"Asymptotic Behavior of n^{th} order Differential Equations;" *M.W. Hudgins,* Saint Leo College, "Dense Partitions in Topological Spaces;" *Steven Cunningham,* Harris Corporation, "Application of Spectral Techniques to Rayleigh Wave Multipathing;" *Alan Wayne,* Pasco Hernando Community College, "The Relative Frequency of Obtuse Triangles;" *Alice Mason,* University of Florida, "Monotone Mappings on n-dimensional Continua;" *David Alon Rose,* St. Leo College, "When Is Exponentiation Commutative?" *Scott Demsky,* University of Miami, "Modular Addition Sets;" *E.P. Miles,* Florida State University, "Mathematical Education Applications of Color Graphics;" *Ben Fusara,* Salisbury State College, "Summer Workshops as an MAA Activity; *Robert Kalin,* Florida State University, "Opportunities for Mathematically Talented High School Students Through the National and State Mu Alpha Theta Math Clubs"; *Edwin G. Laundauer* and *William M. Glasgow,* Naval Nuclear Power School, "Results Involving Differences of Squares;" *Carey Witkov,* Broward Community College, "The Laws of Form;" *Richard M. Ingle,* Computer Sciences Corporation, "Solving Fluid Networks in a Real Time Computer Simulation Model;" *Judith Gersting,* University of Central Florida, "An Alternative to the Classroom Lecture;" *M.S. Jagadish,* University of Miami, "A Simple Proof of the Infinite Product Expansions of Sine and Cosine;" *Peter M. Ryan,* Jacksonville University. "Computer Program for Elementary Statistics Service Courses."

The Association for Women in Mathematics sponsored a session. In addition to a business meeting, the following talks were presented: *Kathleen Timmer,* Jacksonville University, "Groups of Central Type;" *Don Hill,* Florida A&M University, "Women and Mathematics."

The Florida Section each year holds regional meetings so that everyone in the state and the Caribbean can attend meetings without extensive traveling. Well-attended meetings were held at Pensacola, Tallahassee, Gainesville, Suncoast Region, Goldcoast Region, and, for the first time, in Puerto Rico. Detailed reports are available from the secretary.

Saturday, March 14, Chairman Edwin Duda presided at the luncheon-business meeting. Committee reports were presented. Election results were: Professor *Charles Nelson* of University of Florida Chairman-elect; Professors *Marilyn Repsher* of Jacksonville University and *Maurice Nott* of St. Petersburg Junior College, Vice-chairmen.

Frank L. Cleaver, Secretary, Florida Section

OKLAHOMA-ARKANSAS SECTION MEETING

The Oklahoma-Arkansas Section of MAA met March 27-28, 1981, in Oklahoma City, Oklahoma. The attendance was approximately one hundred twenty-five with thirty-eight papers given by faculty and students in the section. The highlights of the meeting were the N.A. Court Lecture given by Dr. *Harold V. Huneke,* University of Oklahoma, and the Invited Address give by Dr. *Leonard Gillman,* University of Texas, Austin.

The following talks completed the program: "The Gauss Code of a Planar Curve, "*Morris L. Marx,* University of Oklahoma; "A Correlation Coefficient Bound," *Franklin Kemp,* Amoco Production Co. Research; "On Spherical Trigonometry," *Naoki Kimura,* University of Arkansas at Fayetteville; "Universally Non-Integrable Functions on an Algebra of Sets," *Michael Keisler,* Arkansas Tech University; "The Application of Integral Transforms to Distribution Problems in Statistics", *Melvin D. Springer,* University of Arkansas at Fayetteville; "A Discussion of Satellite Transfer Orbits," *Rolan Christofferson,* Oklahoma State University; "An Unusual Application of Some Undergraduate Mathematics," *Jack Hamm,* Arkansas Tech University; "On Stability of Systems of Second Order Differential Equations, *Sherwin J. Skar,* Oklahoma State University; "A Short Proof of a Criterion for Weak Convergence of Measures," *Jerry Johnson,* Oklahoma State University; "When is C(X) Flat?" *Robert C. Smith,* University of Arkansas at Fayetteville; "The Inverse Function Theorem, Newton's Method and The Contraction Mapping Principle," *William Ray,* University of Oklahoma; "Inverse Divergences and Riccati Inequalities for Elliptic P.D.E.'s, *Stan Eliason,* University of Oklahoma; "Non-linear Derived Functions," *David C.*

Sutherland, Hendrix College; "Infinite Sums of Derivatives," *Carol Smith*, Hendrix College; "The Exponential Calculus," *Ben Schumacher*, Hendrix College; "Infinite Composition," *Sandra Cousins*, Hendrix College; "General vs. Abstract, a Concrete Example," *Verbal Snook*, Oral Roberts University; "WMG, a Partnership of Teachers and Industrialists in England," *Jeanne Agnew*, Oklahoma State University; "An Algebra Placement Examiniation," *Murray Blose* and *Gerald Goff*, Oklahoma State University; "Teaching Geometry in High School - Some Comments on Preparing Teachers," *Douglas B. Aichele*, Oklahoma State University; "An Example of a Group," *Dennis Bertholf*, Oklahoma State University; "Jacobian Problem - High School Version of Inverse Functions Theorem," Dr. *S. C. Kochari*, University of Oklahoma; "Disjoint Sets in Free Products of Lattices and Lattice Ordered Groups," *Wayne B. Powell*, Oklahoma State University; "Some Results Concerning Borsuk's Hyperspace 2^X_h," *David E. Rowe*, University of Oklahoma; "Weakly 0-Conversion Functions," *Paul E. Long*, University of Arkansas at Fayetteville; "The Answer to an Unsolved Problem in THE MONTHLY," *R.B. Deal*, O.U.H.S.C.; "Stirling Formula from the Central Limit Theorem," *Tetsundo Sekiguchi*, University of Arkansas at Fayetteville; "Generalized Kernels and Their Application in Characterization Problems in Probability, Statistics, and Integral Transforms," *Ignacy Icchak Kotlarski*, Oklahoma State University; "A Bumper Sticker Distribution Model," *Donald L. Patten*, University of Oklahoma; "The Contact Point of Interceptor and Penetrator Missiles," *Pedro E. Adams*, Oklahoma State University; "Introduction to Babylonian Mathematics," *Roy Cowan*, Oklahoma Christian College; "Babylonian Square and Cube Roots," *Bryan Williams*, Oklahoma Christian College; "Babylonian Quadratic Equations," *Kyle Hagar*, Oklahoma Christian College; "A Problem Involving Volume," *Jim Mann*, Oklahoma Christian College; "Graph Theory for Undergraduates," *Paul Duvall*, Oklahoma State University; "Mathematics at the Arkansas Governors School," *Cecil McDermott*, Hendrix College; "A Career Awareness Package for Secondary School Students," *John Jobe*, Oklahoma State University.

MARCH MEETING OF THE NORTHERN CALIFORNIA SECTION

The annual meeting of the Northern California Section was held March 14, 1981, jointly with the Northern California Section of SIAM, at the University of Santa Clara in Santa Clara.

The membership elected *Roy Ryden* from Humboldt State University as vice-chairman and *Leonard Klosinski* from the University of Santa Clara as secretary-treasurer. Succeeding *Henry Osner* as newsletter editor is *William Chinn*.

The following invited papers were presented: "Air Pollution and Health: Acute Effects." *Alice A. Whittemore*, Stanford University; "Optimal Strategies in Sports," *Leonard Gillman*, University of Texas, Austin; "Mathematical Aspects of Computerized Tomography," *F. Alberto Grunbaum*, University of California, Berkeley; "The Relationship Between Number Theory and Analysis," *Paul J. Cohen*, Stanford University.

A luncheon in the Willeman Room at the University of Santa Clara featured a talk by *S. S. Chern* of the University of California, Berkeley on "Mathematics in China: A Recent Assessment." The day's program was concluded with a meeting for department heads and MAA representatives.

SPRING MEETING OF THE NEW JERSEY SECTION

The New Jersey Section of the MAA held its annual spring meeting on Saturday, March 27, 1981 at Seton Hall University, South Orange, New Jersey. Approximately 60 persons attended the meeting which was held in conjunction with MATYCNJ

Prof. *Victor Guillemin* of Massachusetts Institute of Technology gave first a talk on "Geometric Quantization." The second speaker was Prof. *John Saccoman*, of Seton Hall University who spoke on "Some Aspects of the Hahn-Banach Theorem." Next was Prof. *Roe Goodman* of Rutgers University who discussed "Finite Reflection Groups and Integrable Hamiltonian Systems." The final speaker of the morning was Prof. *Ronald Infante* of Seton Hall University who discussed "Sub-Matrices of Orthogonal Matrices."

After lunch there were two more speakers. The first was Prof. *Sunday Jose* of Essex County College whose topic was "The Transition from Remedial to College Level Mathematics." The last speaker was Prof. *Douglas West* of Princeton University who spoke on "Extremal Problems in Graph Theory."

In addition to the speakers, the first New Jersey meeting for Association of Women in Mathematics was held during the section meeting.

MARCH MEETING OF THE WISCONSIN SECTION

The spring meeting of the Wisconsin Section of MAA was held at the University of Wisconsin at La Crosse, March 27-28, 1981, with 74 in attendance.

The invited addresses were: "Mathematics Tomorrow," *Lynn A. Steen*, St. Olaf College; "Unifying Some Theorems of Euclid," *Seymour Schuster*, Carleton College. The following additional talks were given: "Teaching Problem Solving in an Introductory Computer Science Class," *David Riley*, UW-La Crosse; "Mascheroni Constructions," *Orville Bierman*, UW-Eau Claire; "On Polynomial Interpolation in the points of a Geometric Progression," *I.J. Schoenberg*, UW-Madison; "Tesselations and Region-(Space-) Filling Curves (Surfaces)," *Joseph L. Teeters*, UW-Eau Claire; "Numerical Solution of PDE's Describing Fluid Flow in a Square Cavity," *Bill Shay*, UW-Green Bay; "Sex-related Issues in Mathematics Achievement," *Lindsay Anne Tartre*, UW-Madison; "Interval Arithmetic as an Introduction to Numerical Analysis," *Louis B. Rall*, UW-Madison; "A Generalization of Vandermonde's Array," *Dan Kalman*, UW-La Crosse; "Some Notable Women Mathematicians," *Julia Ann Walker*, UW-La Crosse; "Minimal Supplements of a Normal Subgroup," *Roger Erickson*, UW-La Crosse; "Mercator's Projection," *Eli Maor*, UW-Eau Claire; "Micro-Computer Graphics in Precalculus," *Dennis Mich*, Carroll College; "Graphs with Distinct Subgraphs," *David Bange*, UW-La Crosse; "Early Criticism of Symbolic Algebra: the 1830's," *Helena M. Pycior*, UW-Milwaukee.

The film program included "Challenge in the Classroom: the Methods of R.L. Moore;" "Gottingen and New York--Reflections on a Life in Mathematics: Richard Courant;" "Predicting at Random: a Lecture by David Blackwell"--all produced by the MAA.

The program also included a swap session for users of Apple computers.

It was decided not to hold a Fall Workshop in 1981 but to concentrate efforts on a Summer Seminar on computer science for mathematicians in 1982.

CALENDAR OF FUTURE MEETINGS

Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.
Sixty-second Summer Meeting, Toronto, Canada, August 23–25, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, last weekend in April or first weekend in May. Deadline for papers six weeks before meeting.
EASTERN PENNSYLVANIA AND DELAWARE, Villanova University, Villanova, Pennsylvania, November 21, 1981.
FLORIDA, Valencia Community College, Orlando, March 5–6, 1982.
ILLINOIS, first Friday/Saturday in May.
INDIANA, Purdue University, West Lafayette, October 17, 1981.
INTERMOUNTAIN
IOWA, third weekend in April. Deadline for papers February 1.
KANSAS, March or April. Deadline for papers January 1.
KENTUCKY, University of Kentucky, Lexington, April 2–3, 1982.
LOUISIANA–MISSISSIPPI, University of Southwestern Louisiana, Lafayette, February 12–13, 1982.
MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, George Washington University, Washington, D.C., November 14–15, 1981.
METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.
MICHIGAN, first Friday and Saturday in May. Deadline for papers six weeks before meeting.
MISSOURI, University of Missouri, Rolla, April 9–10, 1982.
NEBRASKA, Kearney State College, Kearney, April 2–3, 1982.
NEW JERSEY, Trenton State College, Trenton, fall 1981.
NORTH CENTRAL, Bemidji State University, Bemidji, Minnesota, October 23–24, 1981.
NORTHEASTERN, Saturday before Thanksgiving and third week in June.
NORTHERN CALIFORNIA, first or second Saturday in February.
OHIO, Lorain County Community College, Elyria, October 23–24, 1981.
OKLAHOMA–ARKANSAS, University of Arkansas, Fayetteville, March 25–27, 1982.
PACIFIC NORTHWEST, second Saturday in June. Deadline for papers six weeks before meeting.
ROCKY MOUNTAIN, last weekend in April or first in May. Deadline for papers eight weeks before meeting.
SEAWAY, SUNY, College at Brockport, Brockport, New York, November 6–7, 1981.
SOUTHEASTERN
SOUTHERN CALIFORNIA, University of California, Santa Barbara, November 13–14, 1981.
SOUTHWESTERN, usually in April. Deadline for papers two weeks before meeting.
TEXAS, Friday and Saturday in early April. Deadline for papers March 1.
WISCONSIN, University of Wisconsin, Fond du Lac, late March 1982.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES, Hyatt-Regency Hotel, New Orleans, Louisiana, October 7–11, 1981.
AMERICAN MATHEMATICAL SOCIETY, Cincinnati, Ohio, January 13–16, 1982.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.
ASSOCIATION FOR SYMBOLIC LOGIC, Philadelphia, Pennsylvania, December 1981.
ASSOCIATION FOR WOMEN IN MATHEMATICS
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, Canada, April 14–17, 1982.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Regency Hyatt House, Houston, Texas, October 11–14, 1981.
PI MU EPSILON
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION, Carrousel Inn, Columbus, Ohio, November 5–6, 1981.
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, The Netherland Hilton Hotel, Cincinnati, Ohio, October 26–28, 1981.

Eminent Mathematicians and Mathematical Expositors speak to

STUDENTS and TEACHERS in...

The NEW MATHEMATICAL LIBRARY

An internationally acclaimed paperback series providing·

- stimulating excursions for students beyond traditional school mathematics
- supplementary reading for school and college classrooms
- valuable background reading for teachers
- challenging problems for solvers of all ages from high school competitions in the US and abroad

The New Mathematical Library is published by the MATHEMATICAL ASSOCIATION OF AMERICA. The volumes are paperbound.

For information regarding the price of these publications, please contact The Mathematical Association of America at the address listed below.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven, NML-01

WHAT IS CALCULUS ABOUT? By W. W. Sawyer, NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman, NML-03

GEOMETRIC INEQUALITIES, By N. D. Kazarinoff, NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Contests sponsored by the MAA, NCTM, Mu Alpha Theta, The Society of Actuaries, and the Casualty Actuarial Society. Covers the period 1950-1960. Compiled and with solutions by C. T. Salkind. NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis, NML-06

USES OF INFINITY, by Leo Zippin, NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields, NML-08

CONTINUED FRACTIONS, by C. D. Olds, NML-09

GRAPHS AND THEIR USES, by Oystein Ore, NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eötvos Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport, NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, by A. Aaboe, NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus, NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven, NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs, NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Contests for the period 1961-1965. NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod, NML-18

GEOMETRY REVISITED, by H.S.M. Coxeter, and S. L. Greitzer, NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore, NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields, NML-21

ELEMENTARY CRYPTANALYSIS — A Mathematical Approach. by Abraham Sinkov, NML-22

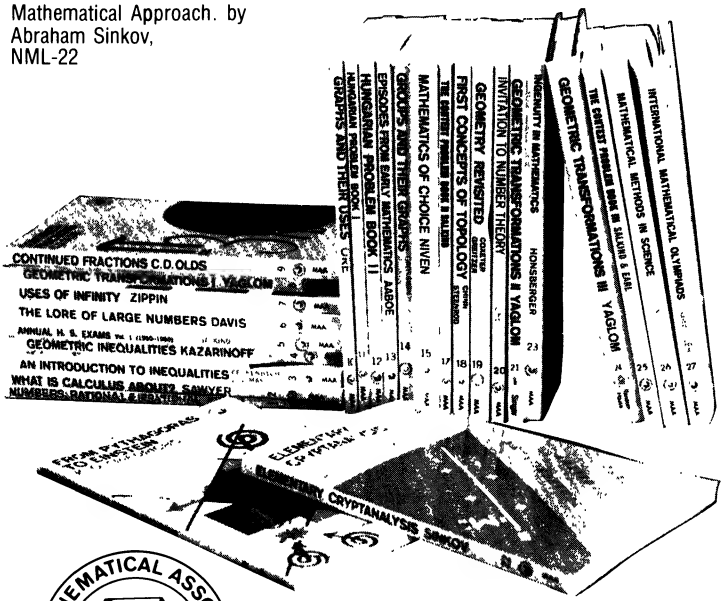
INGENUITY IN MATHEMATICS, by Ross Honsberger, NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer, NML-24

THE CONTEST PROBLEM BOOK III. A continuation of NML-05 and NML-17, containing problems and solutions from the Annual High School Mathematics Contests for the period 1966-1972. NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Pólya, NML-26

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977. Problems, with solutions, from the first nineteen International Mathematical Olympiads. Compiled and with solutions by S. L. Greitzer. NML-27



Send orders to: **The Mathematical Association of America**
1529 Eighteenth St., N.W., Washington, D.C. 20036

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

OCTOBER

THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 8

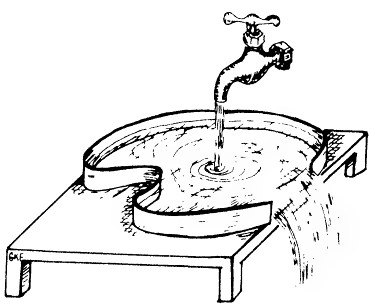
An Experience in Problem Solving

Stability of Populations

Uniformization Theorem

Sonya Kovalevsky

Black Women in Mathematics



Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 — NUMBER 8

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

Annual dues for members of the Association, including a subscription to the American Mathematical Monthly, are \$25 for each of the first two years of membership and \$30 thereafter. Student membership is available with annual dues of \$15. For nonmembers the subscription price is \$36.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

AN EXPERIENCE IN PROBLEM SOLVING

M. KATCHALSKI
Department of Mathematics, The Technion, Haifa, Israel

M. S. KLAMKIN
Department of Mathematics, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

A. LIU
Department of Mathematics, University of Alberta, Edmonton, Alberta T6G 2G1, Canada

1. Introduction. In mathematics, it often happens that we start out with a problem and are then led to consider related ones. This is particularly true in solving problems via transformations [**11**]. Here, we start out with a seemingly innocuous problem, which is essentially to evaluate a definite integral. It is then natural to consider a related extremal problem. This in turn leads to a number of other related extremal problems with different formulations.

We now take you on a small tour in problem solving, which we describe in the Pólya manner (using some headings from [**14**, vol. 2, pp. 36–53]) and starting from a proposal of Kestelman [**10**].

Problem 1 (Calculus). Determine the maximum of

$$\int_0^{2\pi} \cos(m_1 x)\cos(m_2 x)\cdots\cos(m_n x)\,dx$$

where $m_1, m_2, \dots, m_n$ are positive integers.

2. An Auxiliary Problem. First notice that by repeated use of the trigonometric identity

$$2\cos x \cos y = \cos(x+y) + \cos(x-y),$$

we can rewrite the integral as

$$2^{-n}\sum \int_0^{2\pi} \cos x(a_1 m_1 + a_2 m_2 + \cdots + a_n m_n)\,dx$$

where the $a_i = \pm 1$ for $1 \leqslant i \leqslant n$ and the summation is taken over all the 2^n combinations of the values of the a_i. Since

$$\int_0^{2\pi} \cos(n\,x)\,dx = \begin{cases} 2\pi, & \text{if } n = 0 \\ 0, & \text{if } n \text{ is any nonzero integer,} \end{cases}$$

the integral is a maximum if the number of solutions $(a_1, a_2, \dots, a_n)$ to the equation

$$a_1 m_1 + a_2 m_2 + \cdots + a_n m_n = 0$$

is a maximum.

3. A More Ambitious Auxiliary Problem. We are tempted to drop the condition that the m_i are integers.

Problem 2 (Number Theory). If the m_i are positive real numbers and $a_i = \pm 1$ for $1 \leqslant i \leqslant n$, what is the maximum number of solutions $\{a_i\}$ that the equation

$$a_1 m_1 + a_2 m_2 + \cdots + a_n m_n = 0$$

can have?

Since this does not seem to give us any immediate help, we continue our search for related auxiliary problems.

Partition the set $\{1,2,\dots,n\}$ into two subsets A and B, putting $i \in A$ if $a_i = 1$ and $i \in B$ if

$a_i = -1$. The equation in Problem 2 then becomes

$$\sum_{i \in A} m_i = \sum_{i \in B} m_i.$$

If we regard $S = \{m_1, m_2, \ldots, m_n\}$ as a set of weights, this means a partition of S into two balancing subsets. So here we have an auxiliary problem equivalent to Problem 2.

Problem 3 (Combinatorics). A set of n weights is partitioned into two sets so that they balance each other on a two-pan balance. In at most how many ways can this be done?

4. A Stimulating Influence. Problem 3 has made things clearer and more accessible, so that we can begin in earnest to attempt a solution.

Let $N(n)$ denote the maximum number of ways of partitioning n weights into two balancing subsets. Intuitively, we expect this maximum number to occur when the weights are as equal as possible to one another. In particular, if $n = 2k$ or $2k + 1$ and we choose $m_1 = m_2 = \cdots = m_k$ (and $m_{2k+1} = 2m_1$ for odd n), we have

$$N(2k) \geqslant \binom{2k}{k}, \qquad N(2k+1) \geqslant 2\binom{2k}{k-1}.$$

But to establish equality by proving the reverse inequalities

$$N(2k) \leqslant \binom{2k}{k} \tag{1}$$

and

$$N(2k+1) \leqslant 2\binom{2k}{k-1} \tag{2}$$

is not a trivial matter. We have a situation, familiar to many combinatorists, sometimes called "one-legged induction." The simultaneous induction would work readily for (1), but not for (2). Other attempts also failed and each time it was the odd case that was the stumbling block.

It seems that a direct attack is not the proper approach at this stage. However, the formulation of Problem 3 stimulates us to make the following key observation:

$N(n)$ is the number of subsets A of the set S of weights such that the total weight of A is one-half that of S. It is clear that its complement $\bar{A}$ has the same property but not any supersets or subsets of A.

This is a breakthrough that we have been looking for. It enables us to formulate the following abstraction of Problem 3.

Problem 4 (Set Theory). Let S be a set of n elements and $\mathfrak{F}$ a family of subsets of S such that
(i) $\bar{A} \in \mathfrak{F}$ for all $A \in \mathfrak{F}$,
(ii) $A \not\subset B$ for all $A, B \in \mathfrak{F}$.
What is the maximum size of $\mathfrak{F}$?

It is clear that the maximum size of $\mathfrak{F}$ is equal to $N(n)$, which we have defined previously, and Problem 4 is a more ambitious version of Problem 3. This formulation brings immediately to mind the following classical result.

SPERNER'S THEOREM [17]. *Let S be a set of n elements and $\mathfrak{F}$ a family of subsets of S such that $A \not\subset B$ for all $A, B \in \mathfrak{F}$. Then*

$$|\mathfrak{F}| \leqslant \binom{n}{[n/2]}.$$

5. Removing a Clause. Sperner's theorem is just Problem 4 with the condition (i) left out; so it is yet another step up the scale of ambition. However, it does not give us the complete solution to Problem 4, as we will now see.

Inequality (1) is an immediate consequence of Sperner's theorem; but for $n = 2k + 1$, we get

$$N(2k+1) \leqslant \binom{2k+1}{k}.$$

Not quite enough! No great surprise because the odd case has given us trouble before.

We now consider the case of $n = 2k + 1$ of Problem 4 in greater detail. Here S is a set of $2k + 1$ elements and $\mathfrak{F}$ is a family of subsets of S satisfying conditions (i) and (ii). We need only to show that

$$|\mathfrak{F}| \leqslant 2\binom{2k}{k-1}.$$

Now a second classical result appears on the scene.

THE ERDŐS-KO-RADO THEOREM **[4]**. *Let S be a set of n elements and $\mathfrak{F}$ a family of subsets of S such that*

(a) $|A| = k$ *for all $A \in \mathfrak{F}$, where $k \leqslant n/2$ is fixed,*

(b) $A \cap B \neq 0$ *for all $A, B \in \mathfrak{F}$.*

Then,

$$|\mathfrak{F}| \leqslant \binom{n-1}{k-1}.$$

This conclusion is what we are after, but does our family $\mathfrak{A}$ satisfy the hypothesis? Unfortunately, condition (a) does not necessarily hold for $\mathfrak{A}$, and there is some doubt concerning condition (b).

We first remove this doubt. Let A and B be subsets of $\mathfrak{A}$ and suppose $A \cap B = \varnothing$. Then, $A \subset \bar{B}$. By condition (i), $\bar{B}$ as well as A belongs to the family $\mathfrak{F}$ in Problem 4, and this contradicts condition (ii). Hence, we must have $A \cap B \neq \varnothing$ and $\mathfrak{A}$ does satisfy condition (b).

To overcome the difficulty that $\mathfrak{A}$ may not satisfy condition (a), we digress to give Sperner's proof of his theorem **[17]**. We are looking for methodological help.

Proof of Sperner's Theorem. Let $\mathfrak{A}_r = \{A \in \mathfrak{F} \,|\, |A| = r\}$ for $0 \leqslant r \leqslant n$. If $\mathfrak{A}_r = \varnothing$ except for $r = [n/2]$, the theorem will follow at once. Assuming the contrary, suppose there is a least integer $t < [n/2]$ for which $\mathfrak{A}_t \neq \varnothing$. We shall construct a new family $\mathfrak{F}_{t+1}$ which satisfies the hypothesis of the theorem, contains no subsets of size t or less, and has cardinality no less than that of $\mathfrak{F}$.

Define $B_{t+1} = \{B \subset S \,|\, |B| = t+1$ and $B \supset A$ for some $A \in \mathfrak{A}_t\}$ and $\mathfrak{F}_{t+1} = (\mathfrak{F}_t - \mathfrak{A}_t) \cup B_{t+1}$. Now every $B \in B_{t+1}$ contains at most $t + 1$ A's in $\mathfrak{A}_t$. Hence,

$$(n-t)|\mathfrak{A}_t| \leqslant (t+1)|B_{t+1}|$$

or

$$|B_{t+1}| \geqslant \frac{n-t}{t+1}|\mathfrak{A}| > |\mathfrak{A}|$$

as $t < [n/2]$. Since clearly $\mathfrak{F} \cap B_{t+1} = \varnothing$, we have $|\mathfrak{F}_{t+1}| > |\mathfrak{F}|$, and it is easy to see that $\mathfrak{F}_{t+1}$ has all the desired properties as well.

By repeating this process, we can eliminate by substitution all subsets in $\mathfrak{F}$ of size less than $[n/2]$. Similar action can be initiated against subsets of size greater than $[n/2]$. Hence, the theorem holds. □

We now return to consider the family $\mathfrak{A}$ in the odd case of Problem 4. If every subset in $\mathfrak{A}$ is of

size k, we can appeal directly to the Erdős-Ko-Rado theorem to obtain

$$|\mathcal{C}| \leqslant \binom{2k}{k-1}.$$

If there are subsets of size less than k, the argument in the proof above shows that $\mathcal{C}$ will not be any larger. In either case, inequality (2) holds.

6. Sperner's Theorem Revisited. So that our exposition will be self-contained, we give a proof of the Erdős-Ko-Rado theorem. On the way we are led to two more related extremal problems.

In the spirit of our discussion, we digress once more to seek methodological help. We give an alternative proof of Sperner's theorem and it is very instructive to compare the two different approaches.

First we give a definition. A maximal chain of subsets of a set S of n elements is a chain $\varnothing = S_0 \subset S_1 \subset \cdots \subset S_n = S$ with $S_i = i$ for $0 \leqslant i \leqslant n$. It is not difficult to see that S has $n!$ maximal chains.

Alternate Proof of Sperner's Theorem. Let F_i, $1 \leqslant i \leqslant m$, be the subsets of S in $\mathcal{F}$ and let $\mathcal{C}_j$, $1 \leqslant j \leqslant n!$, be the maximal chains of S. Construct an m by $n!$ matrix M where

$$M(i,j) = \begin{cases} 1 & \text{if } F_i \in \mathcal{C}_j, \\ 0 & \text{otherwise.} \end{cases}$$

Now the sum of the ith row is $|F_i|!(n-|F_i|)!$, while the sum of each column is at most 1, as $\mathcal{F}$ can contain at most one subset from each chain. Hence,

$$\sum_{j=1}^{n!} 1 \geqslant \sum_{i=1}^{m} |F_i|!\,(n-|F_i|)!$$

or

$$1 \geqslant \sum_{i=1}^{m} 1 \Big/ \binom{n}{F_i} \geqslant m \Big/ \binom{n}{[n/2]}.$$

It follows that

$$|\mathcal{F}| = m \leqslant \binom{n}{[n/2]}. \qquad \square$$

This elegant proof is due to Lubell [**12**], and we shall extend his idea to give an alternative proof of the Erdős-Ko-Rado theorem. Before we do this, we pause to add to our problem list.

Problem 5 (Linear Programming). Maximize $(x_1 + x_2 + \cdots + x_8)$ subject to $x_i \geqslant 0$ for $1 \leqslant i \leqslant 8$ and

$$\begin{aligned} x_1 + x_2 + x_5 + x_8 &\leqslant 1, \\ x_1 + x_2 + x_6 + x_8 &\leqslant 1, \\ x_1 + x_3 + x_5 + x_8 &\leqslant 1, \\ x_1 + x_3 + x_7 + x_8 &\leqslant 1, \\ x_1 + x_4 + x_6 + x_8 &\leqslant 1, \\ x_1 + x_4 + x_7 + x_8 &\leqslant 1. \end{aligned}$$

This kind of problem can be solved by standard techniques, but these may be tedious, and in general the solutions are not necessarily in integers. However, the problem is readily solved if one recognizes (see [**2**]) that we have in essence Lubell's proof of Sperner's theorem for $n = 3$. Add the

six inequalities to give

$$6 \geqslant 6x_1 + 2(x_2 + x_3 + \cdots + x_7) + 6x_8 \geqslant 2(x_1 + x_2 + \cdots + x_8)$$

or

$$x_1 + x_2 + \cdots + x_8 \leqslant 3.$$

Since this can be attained by letting $x_2 = x_3 = x_4 = 1$ and $x_i = 0$ otherwise, the problem is solved.

It is to be pointed out that if Problem 5 is posed as an Integer Programming problem, then it is a bilateral reduction of the corresponding case of Sperner's theorem. As it stands, it is more ambitious.

7. Discrete arcs. While seeking an alternative proof of the Erdős-Ko-Rado theorem, we were pleasantly surprised to receive help from our experience in solving a problem dealing with quite a different topic. Before we formulate this more remote auxiliary problem, we again need some definitions.

A *discrete circle* of order n is a set of n ordered points $a_1, a_2, \ldots, a_n$. An *arc* $[a_k, a_j]$ on this circle is the set of points $a_i, a_{i+1}, \ldots, a_j$ with a_{n+i} being identified with a_i. The length of an arc A, denoted by $|A|$, is the number of points it contains.

Katchalski and Liu [7] have considered intersection properties of discrete arcs which lead to a "circular" version of the Erdős-Ko-Rado theorem.

Problem 6 (Geometry). Let $\mathcal{F}$ be a family of arcs on a discrete circle of order n such that
(a) $|A| = k$ for all $A \in \mathcal{F}$ where $k \leqslant n/2$ is fixed.
(b) $A \cap B \neq \varnothing$ for all $A, B \in \mathcal{F}$.
What is the maximum size of $\mathcal{F}$?

We may assume that the arc $[a_1, a_k]$ belongs to $\mathcal{F}$. By conditions (a) and (b), the other arcs in $\mathcal{F}$ are among the following:

$$\begin{array}{ll} [a_{n-k+2}, a_1], & [a_2, a_{k+1}], \\ [a_{n-k+3}, a_2], & [a_3, a_{k+2}], \\ \vdots & \vdots \\ [a_n, a_{k-1}], & [a_k, a_{2k-1}]. \end{array}$$

Since $k \leqslant n/2$, the two arcs on each row in the list above are disjoint; and, by condition (b), $\mathcal{F}$ can contain at most one from each pair. Hence, $|\mathcal{F}| \leqslant k$, and this maximum is attained by choosing the first arc of each pair.

Now that we have solved Problem 6, we are ready to give our alternative proof, which is adapted from Katona [9].

Proof of the Erdős-Ko-Rado Theorem. Let S be regarded as a discrete circle. Also, let F_i, $1 \leqslant i \leqslant m$, be the subsets of S in $\mathcal{F}$ and P_j, $1 \leqslant j \leqslant n!$, be the permutations of the elements of S. Construct an m by $n!$ matrix M where

$$M(i,j) = \begin{cases} 1 & \text{if } P_j(F_i) \text{ is an arc,} \\ 0 & \text{otherwise.} \end{cases}$$

The sum of each row of M is $n(k!)(n-k)!$, there being n arcs of length k. On the other hand, for $1 \leqslant j \leqslant n!$, let $\mathcal{F}_i = \{P_j(F_i) | M(i,j) = 1\}$. Clearly, $\mathcal{F}_j$ satisfies the hypotheses of Problem 6. Thus, the sum of column j is at most k for all j. Finally,

$$k(n!) \geqslant mn(k!)(n-k)!$$

or

$$|\mathscr{F}| = m \leqslant \binom{n-1}{k-1}. \qquad \square$$

For advanced treatments of this subject, see [**5**], [**6**], and [**8**].

8. *Respice finem.* We thought we were unlucky when the one-legged induction failed in Section 4. But now that we have worked through five auxiliary problems and familiarized ourselves with two classical results, we see how lucky we have been!

Problem 3 (see [**1**] and [**16**] for other abstractions) suggests another natural question. In at most how many ways can a set of n weights be partitioned into more than two subsets so that they balance one another pairwise? A generalization of Sperner's theorem (see [**13**]) may help, but so far this problem is still unsolved.

Oh! We nearly forgot!

$$\max \int_0^{2\pi} \cos(m_1 x)\cos(m_2 x)\cdots\cos(m_n x)\,dx = \begin{cases} \dbinom{2k}{k} & \text{if } n = 2k, \\ 2\dbinom{2k}{k-1} & \text{if } n = 2k+1. \end{cases}$$

References

1. B. Bollobás, Sperner's systems consisting of pairs of complementary subsets, J. Comb. Theory Ser. A, 15 (1973) 363–366.

2. V. Chvatal, Some linear programming aspects of combinatorics, Algebraic Aspects of Combinatorics, Utilitas Math. Publ. Inc., Winnipeg, 1975, pp. 2–30.

3. P. Erdős, On a lemma of Littlewood and Offord, Bull. Amer. Math. Soc. 51 (1945) 898–902.

4. P. Erdős, Chao Ko, and R. Rado, Intersection theorem for systems of finite sets, Quart. J. Math. Oxford Ser. 2, 12 (1961) 313–320.

5. P. Erdős and D. Kleitman, Extremal problems among subsets of a set, Discrete Math., 8 (1974) 281–294.

6. C. Greene and D. Kleitman, Proof techniques in the theory of finite sets, in Studies in Combinatorics, MAA Studies, vol. 17, ed. G. Rota, 1978, pp. 22–79.

7. M. Katchalski and A. Liu, Arcs on the circle and p-tuplets on the line, Discrete Math., 27 (1979) 59–69.

8. G. Katona, Extremal problems for hypergraphs, Combinatorics, Reidel, Boston, 1974, 215–244.

9. ______, A simple proof of the Erdős-Ko-Rado Theorem, J. Comb. Theory Ser. B, 13 (1972) 183–184.

10. H. Kestelman, Problem 1033, Math. Mag., 51 (1978) 127; Z. Motteler, Solution to Problem 1033, Math. Mag., 52 (1979) 182–183.

11. M. S. Klamkin, D. J. Newman, The philosophy and applications of transform theory, SIAM Review, 3 (1961) 10–36.

12. D. Lubell, A short proof of Sperner's Theorem, J. Comb. Theory, 1 (1966) 299.

13. G. Pólya, How to Solve It, Doubleday-Anchor, New York, 1957.

14. ______, Mathematical Discovery I, II. Wiley, New York, 1962.

15. ______, Mathematics and Plausible Reasoning I, II. Princeton University Press, Princeton, N.J., 1954.

16. J. Schönheim, On a problem of Purdy related to Sperner systems, Canad. Math. Bull., 17 (1974) 135–136.

17. E. Sperner, Ein Satz über Untermengen einer endlichen Menge, Math Z., 27 (1928) 544–548.

MISCELLANEA

60. One might just as well maintain that continuous but not differentiable functions are unimportant because they are artificially constructed— a term which I suppose means that they do not present themselves when unasked for.— P. E. B. Jourdain, *The Philosophy of Mr. B*rtr*nd R*ss*ll*, London, 1918, p. 76.

WHY A POPULATION CONVERGES TO STABILITY

W. B. ARTHUR
International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

1. Introduction. Central to both mathematical demography and its sister subject, population biology, is a single, fundamental theorem: If the reproductive and the survival age-patterns of a population remain unchanged over time, its age composition, no matter what its initial shape, will converge over time to a fixed and persistent form. In brief, when demographic behavior remains unchanged, the population, it is said, converges to stability. This is the Strong Ergodic Theorem of Demography.

It is this theorem that makes *stable population theory* possible. Normally there is no direct connection between reproductive and survival behavior and the age composition. But where reproductive and survival age-patterns do not fluctuate greatly over time the theorem tells us that a known and fixed age composition becomes associated with them. We can use this relationship between individual life-cycle behavior and the population age composition widely: in demographic analyses, in population projections, and in the estimation of vital rates.

At present, there are two basic ways to prove population convergence to stability, depending on whether population dynamics are described in discrete time, or in continuous time. For the discrete case, where the dynamics are described by a transition matrix that maps the age-group vector at time t into that at $t + 1$, proof amounts to invoking the Perron-Frobenius theorem for nonnegative primitive matrices. This is the principle behind the arguments of Leslie [**5**], Lopez [**6**], Parlett [**10**], and many others. For the continuous case, where the dynamics are described by a renewal equation that equates present births with the reproduction of those born in the past, proof amounts to solving the Lotka-Volterra integral equation and studying the behavior of its solution terms. This is the principle behind the arguments of Lotka and Sharpe [**7**], Lotka [**8**], Coale [**1**], and, again, many others. Some of the modern papers, Cohen [**2**], for example, extend the theorem to a stochastic version, but the underlying principles remain largely the same.

For the professional, all this is fine. The theorem has been proved two ways, and neither proof is inordinately difficult. For others—undergraduates, say—things are less satisfactory. Stable population analysis has become indispensable in both demography and population biology (see Keyfitz [4]). Yet the theory that underlies it is not readily accessible to undergraduates unless they have been exposed to either positive matrix theory or Tauberian theory. Even if they have, neither form of proof offers much in the way of direct insight into *why* the age composition should converge to a fixed and final form. The mechanism remains within the "borrowed" theorems, and becomes difficult to see. The only proof I know of that does not appeal to outside theorems was published by Lotka [**9**] in 1922. Lotka sandwiched the initial age composition between two boundary curves that close in over time, eventually coinciding to trap the age composition within a fixed shape. But here the logic is loose, and the mechanism forcing convergence is still partially hidden from view.

In this expository paper, I put forward a new proof of the convergence of the age structure, one that is self-contained, that brings the mechanism behind convergence into full view, and that uses elementary operations only. The idea is simple. Looked at directly, the dynamics of the age composition say little to our normal intuition. Looked at from a slightly different angle, however, we see that population dynamics decompose into a growth process and a smoothing or averaging process over the generations—processes comfortable to our intuition. It is this smoothing and resmoothing of the initial birth sequence that forces the age composition toward a limiting form.

W. Brian Arthur is Research Scholar with the International Institute for Applied Systems Analysis at Laxenburg, Austria. He received his Ph.D. in 1973 from the University of California, Berkeley, in Operations Research, under Stuart E. Dreyfus. His main research interests are in mathematical demography, mathematical economics, and optimization theory.—*Editors*

2. The Problem. Consider a one-sex population that has survival and reproductive age patterns constant in time, that has no in- or out-migration, and that evolves over discrete time according to the dynamics

$$B_t = \sum_x B_{t-x} p_x m_x. \tag{1}$$

B_t is the number of births in year t, m_x is the proportion of those at age x who reproduce at that age, and p_x is the proportion of those born who survive until age x. Present births, in other words, are the sum of births born to individuals at reproductive ages who still survive. Summation is understood as taken over the age groups 1 to M (where M is an upper limit to childbearing, 45 years, say, for humans). And the numbers in the initial *generation*, $B_{-M}, \ldots, B_{-1}$, are assumed given. The birth variable B and reproductive rates m, of course, are nonnegative; the survival rates p are strictly positive.

This simple model of population dynamics, due to Lotka, applies to many animal species. For better precision one sex only is studied; the other may have different vital age patterns. This causes no great problems: the reproduction of one sex determines the births of the other and so the two-sex population can easily be constructed. (It is assumed however that there are enough members of the other sex not to alter the reproductive and survival patterns of the sex we are studying.)

The age composition, or proportion of the population at age a at time t, is given by

$$c_{a,t} = \frac{B_{t-a} p_a}{\sum_x B_{t-x} p_x}, \tag{2}$$

the numbers at age a, divided by the total population. Summation here is understood as taken over ages 1 to N, where N is an upper bound to the length of life. We want to show that, under certain conditions yet to be specified, the time-varying function $c_{a,t}$ converges to a limiting constant function c_a^*.

Two observations will help the argument. First, note that it is enough to show that B_t converges to an exponential form, $B_t \to B^* e^{rt}$, where B^* and r are constant. For if this is true it follows by substituting for B_{t-a} and B_{t-x} in (2) that the age composition becomes fixed and unchanging with time:

$$c_{a,t} \to \frac{e^{-ra} p_a}{\sum_x e^{-rx} p_x} = c_a^*. \tag{3}$$

We can therefore confine our attention to why the birth sequence, B_t, should become exponential. (To ease comparison with the literature, I talk about "exponential" rather than geometric growth. A parameter λ may be substituted for e^{-r} if preferred.) Second, note that convergence to an exponential form—a moving target—is hard to prove; but convergence to a fixed value is easier. I will therefore normalize or redefine the problem to one of convergence to a fixed value.

3. Smoothing Process. Begin with the dynamics

$$B_t = \sum_x B_{t-x} p_x m_x, \tag{4}$$

and divide both sides by e^{rt},

$$\frac{B_t}{e^{rt}} = \sum_x \frac{B_{t-x}}{e^{rt}} \frac{e^{-rx}}{e^{-rx}} p_x m_x. \tag{5}$$

Renaming Be^{-rt} to be the variable $\hat{B}_t$—the "growth-corrected" birth sequence—the new, but equivalent, dynamics become

$$\hat{B}_t = \sum_x \hat{B}_{t-x} e^{-rx} p_x m_x. \tag{6}$$

I will speak somewhat loosely of $\hat{B}$ in what follows as "births," remembering though that these "births" differ from real births by an exponential factor.

We now need only show that for some value of r, $\hat{B}$ eventually becomes constant over time. Allowing ourselves some foresight, we choose r to be the unique real root of the equation

$$1 = \sum_x e^{-rx} p_x m_x. \tag{7}$$

(It is easy to check that (7) has a single real root. For r on the real line the function $1 - \Sigma_x e^{-rx} p_x m_x$ is monotonic increasing from $-\infty$ to $+1$. It therefore attains zero at one and only one value of r.) Finally, renaming $e^{-rx} p_x m_x$ as ψ_x, we may write the new but equivalent dynamics as

$$\hat{B}_t = \sum_x \hat{B}_{t-x} \psi_x \tag{8}$$

where, by virtue of (7),

$$1 = \sum_x \psi_x. \tag{9}$$

The original dynamics have been changed but little; B_t has merely been normalized to the new variable $\hat{B}_t$. Notice though, in the new system for $\hat{B}_t$, the coefficients ψ_x sum to one—ψ is a weighting function. The new dynamics therefore describe a continual smoothing process: $\hat{B}_t$ is the weighted average of the M immediate past values of $\hat{B}$; $\hat{B}_{t+1}$ is the weighted average of $\hat{B}_t$ and the $M-1$ immediate past values of $\hat{B}_t$; $\hat{B}_{t+2}$ the weighted average of $\hat{B}_{t+1}$, $\hat{B}$, and the $M-2$ immediate past values of $\hat{B}_t$. And so on. This constant averaging, then averaging of the averages, we would suspect, will converge $\hat{B}$ to a fixed value B^* (as in Fig. 1), and equivalently will converge B to the exponential form $B^* e^{rt}$. Why?

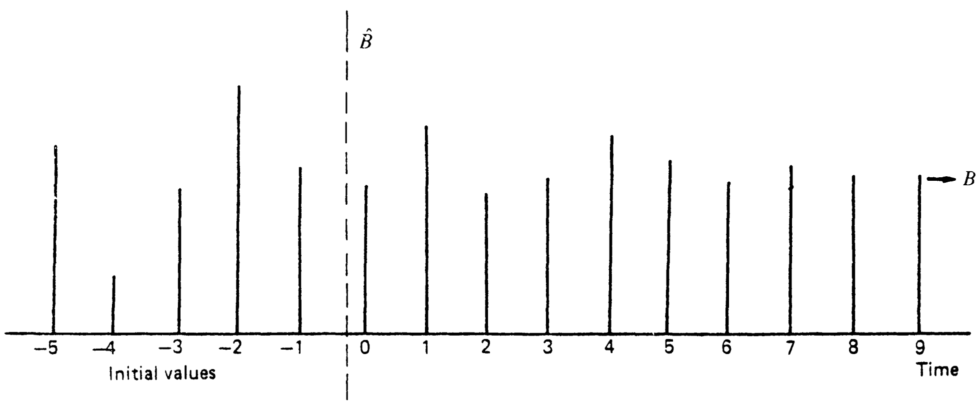


FIG. 1.

The reason is easy to sketch when all ψ_x are strictly positive (strictly greater than ε say). If all the $\hat{B}$-values are identical to begin with, then they are perpetuated at this value and the limit is reached. If not, denote the maximum value of the initial birth values as $\bar{s}$, the minimum as $\underline{s}$, the difference being γ. The initial given values therefore lie within a "spread" of γ units. Without further information on the initial sequence we can say that $\hat{B}_0$ must be strictly smaller than $\bar{s} - \varepsilon\gamma$, for the worst case would have all values equal to the maximum except one that is smaller by at least γ units. It will therefore cause $\hat{B}_0$ to fall short of the maximum by at least $\varepsilon\gamma$ units.

Similarly we can argue that $\hat{B}_0$ must exceed the minimum value $\underline{s}$ by at least $\varepsilon\gamma$ units. Therefore $\hat{B}_0$ will lie strictly inside the initial spread of birth values—inside by a fixed factor $1 - 2\varepsilon$. The same argument applies all the more so to $\hat{B}_1$, and again to $\hat{B}_2$, and so on until $\hat{B}_{M-1}$. The spread of the entire new generation of $\hat{B}$ values therefore lies strictly within that of the old one, and by a

specified uniform factor. Repeating the argument over the generations, the generational spread in $\hat{B}$ diminishes geometrically to zero. $\hat{B}_t$ therefore converges to a fixed value B^*, B_t to exponential growth $B^* e^{rt}$, and the age composition to the fixed form in (3).

Noting that ψ_x is positive or zero depending on whether the reproductive rate m_x is positive or zero, we may summarize our findings up to this point as

THEOREM 1. *A population that has strictly positive reproductive rates at all ages and unchanging reproductive and survival age patterns converges to stability.*

So far so good. But what of real populations where no reproduction takes place at certain ages, where some of the ψ_x values will be zero? Will these populations always converge? The answer is no. Consider the four-age-group population in Fig. 2, with $\psi_1 = \psi_3 = 0$, and $\psi_2 = \psi_4 = \frac{1}{2}$. Childbearing occurs only in the second and fourth age-groups. The size of birth cohorts in this population will oscillate indefinitely. Here the smoothing process does not smooth: the population fails to converge. To lay down conditions for populations that do not reproduce at all ages we need to look at smoothing and convergence more closely.

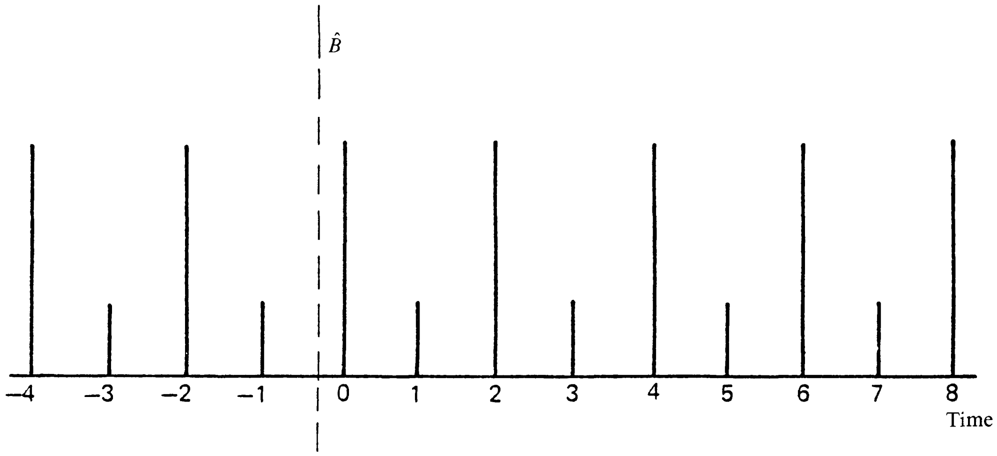


FIG. 2.

4. **Convergence—A Closer Look.**[1] In general, assume that some, or several, of the ψ values are zero. The value $\hat{B}_0$ then depends directly on only certain of the original $\hat{B}$ values. Similarly, $\hat{B}_1$ depends directly on the neighbors of these values. For the system

$$\hat{B}_t = \hat{B}_{t-3}\psi_3 + \hat{B}_{t-4}\psi_4, \quad \psi_1, \psi_2 = 0,$$

we can graph this dependence as in Fig. 3a, picturing each birth cohort as a point, with a directed arrow drawn between them if dependent. The graph extends indefinitely downward. Notice though that while $\hat{B}_0$ depends directly on only two of the initial values, $\hat{B}_4$ depends on three of them, and $\hat{B}_8$ on all four of them. If we so chose, we could therefore write the dynamics with present $\hat{B}$ values specifying $\hat{B}$ eight steps ahead:

$$\hat{B}_{t+8} = \hat{B}_{t-1}\psi_1' + \hat{B}_{t-2}\psi_2' + \hat{B}_{t-3}\psi_3' + \hat{B}_{t-4}\psi_4'.$$

This process, with new weights ψ', describes the evolution of $\hat{B}$ perfectly well; moreover, it remains a smoothing process, as we can see by following the weights backward from $\hat{B}_8$: they divide up but continue to sum to one. Most important, it is a function of *all* the initial values and is strictly positive in all its coefficients. We could therefore apply the convergence argument above, showing that the spread in any four consecutive values must be reduced eight steps ahead by a fixed factor. Taking $\hat{B}$ values now twelve at a time (the original initial four plus the

intervening eight), the spread of each twelve must as before reduce geometrically; $\hat{B}$ converges, even though we started with some ψ weights as zero. Generalizing this argument we can conclude that for a smoothing process: *If there exists a point in the graph that depends on* all *initial points, the process will converge.*

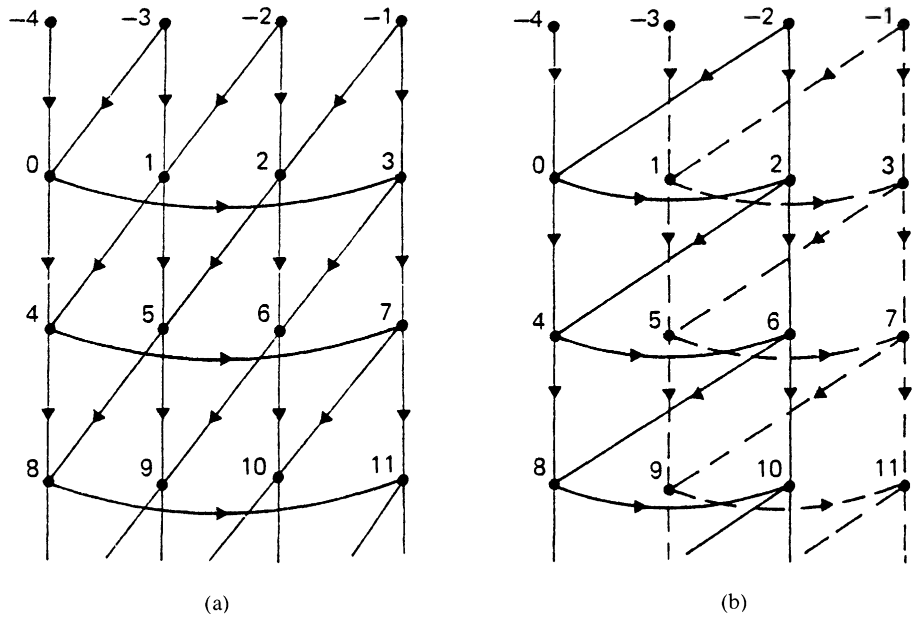


FIG. 3.

What then went wrong with the case where ψ_1 and ψ_3 were zero? Forming its graph (Fig. 3b) we see there is no future $\hat{B}$ value that is a function of all the original given values. Even-indexed $\hat{B}$'s depend on even-indexed $\hat{B}$'s; odd ones depend on odd ones. Here two separate but identical processes are going on: the even process never "sees" the initial values of the odd process and vice versa. Both processes iterate their initial values to a limit, but there is an even limit and an odd limit. The process oscillates indefinitely between them.

We can now deduce some general principles, in each case assuming a population whose reproductive and survival patterns remain unchanged over time, and constructing the smoothing process graph that goes with it. Sufficient for most purposes is a straightforward result:

THEOREM 2. *A population that reproduces at at least two consecutive ages converges to stability.*

In this case any $\hat{B}$ value sufficiently in the future connects in the graph to points that progressively fan out backward to include all initial points. Smoothing takes place and the population converges. This theorem is already enough to cover human populations and those of animal species that reproduce over a block of consecutive ages. But we can sharpen this result, if we want.

THEOREM 3. *A population that reproduces at at least two ages p and q which are relatively prime converges to stability.*

We need to show that some $\hat{B}_K$ in the graph "reaches back" in multiples of p and q to all initial points. This amounts to showing that for $k = 1,\ldots,M$, for some K large enough, we can always find integers a and b so that $ap + bq = K + k$. I leave the details to a note.[2] This theorem is of more mathematical than biological interest. I do not know of any species that satisfies Theorem 3 but not Theorem 2, though some may exist. The converse of Theorem 3 is more interesting, for it applies, sometimes with a vengeance, to populations that reproduce at a single age.

THEOREM 4. *A population that reproduces at ages which have a greatest common divisor* (*greater than* 1) *does not necessarily converge to stability.*

Take the greatest common divisor d and label all points in the graph modulo this number. We then see that the process decouples into d separate, but identical, processes. Each of these processes, considered on its own, can be relabeled with indices at least two of which are relatively prime; each therefore converges to its own limit. (It may happen, of course, that some or all of these limits have the same value.) The overall smoothing process cycles between these limits, with period d. This means that the birth sequence oscillates around its growth path; the age composition, in turn, fails to converge.

Populations which converge tend to settle down again in their birth sequence, sometimes quickly, sometimes slowly, if subjected to an environmental influence that disturbs their numbers. Nonconvergent populations, on the other hand, are prone to booms and busts; their dynamics do not easily "forget" past disturbances; they are not ergodic.

5. The Limiting Coefficient. One question remains. For populations that converge, how can we determine B^*, the limiting coefficient of the exponential birth sequence? One possibility is to look for a quantity that is invariant, that is carried along unchanged over the generations. Such a quantity would enable us to relate B^* at the end of the process to the $\hat{B}$ values at the beginning. Now, each generation at any time can donate to the future a certain number of direct descendants. On these direct descendants all future population must be built—they are the system's "reproductive potential" or "reproductive value" as it were. We might suspect this reproductive potential, *in the growth-corrected dynamics* we have defined, to be invariant. A little algebra shows that this turns out to be the case.

At time t, age-groups $\hat{B}_{t-M},\ldots,\hat{B}_{t-1}$, taken together, contribute V_t direct descendants to the future—to the period from t onward:

$$V_t = \hat{B}_{t-M}\psi_M + \hat{B}_{t-M+1}(\psi_{M-1} + \psi_M) + \cdots + \hat{B}_{t-1}(\psi_1 + \psi_2 + \cdots + \psi_M). \tag{10}$$

Similarly age-groups $\hat{B}_{t-M+1},\ldots,\hat{B}_t$ contribute V_{t+1} to the period from $t+1$ onward:

$$V_{t+1} = \hat{B}_{t-M+1}\psi_M + \hat{B}_{t-M+2}(\psi_{M-1} + \psi_M) + \cdots + \hat{B}_t(\psi_1 + \psi_2 + \cdots + \psi_M).$$

Noting that the coefficient of $\hat{B}_t$ is one, and using (8) to replace $\hat{B}_t$, we find

$$V_{t+1} = \hat{B}_{t-M+1}\psi_M + \hat{B}_{t-M+2}(\psi_{M-1} + \psi_M) + \cdots + \hat{B}_{t-1}\psi_1 + \cdots + \hat{B}_{t-M+1}\psi_{M-1} + \hat{B}_{t-M}\psi_M. \tag{11}$$

Equations (10) with (11), written out fully, can be matched term for term. Therefore V_t equals V_{t+1}; V_t is indeed an invariant quantity V.

At the start

$$V_0 = V = \hat{B}_{-M}\psi_M + \hat{B}_{-M+1}(\psi_{M-1} + \psi_M) + \cdots + \hat{B}_{-1}(\psi_1 + \psi_2 + \cdots + \psi_M). \tag{12}$$

And in the limit

$$\begin{aligned} V &= B^*\psi_M + B^*(\psi_{M-1} + \psi_M) + \cdots + B^*(\psi_1 + \cdots + \psi_M) \\ V &= B^*(\psi_1 + 2\psi_2 + \cdots + M\psi_M). \end{aligned} \tag{13}$$

Since ψ is the distribution of childbearing in the population, the coefficient of B^* is the mean age of childbearing, denoted A_m. Putting (12) and (13) together yields the result we seek:

$$\begin{aligned} B^* &= \frac{1}{A_m}\left\{\hat{B}_{-M}\psi_M + \hat{B}_{-M+1}(\psi_{M-1} + \psi_M) + \cdots + \hat{B}_{-1}(\psi_1 + \cdots + \psi_M)\right\} \\ &= \frac{1}{A_m}\left\{B_{-M}e^{rM}\psi_M + B_{-M+1}e^{r(M-1)}(\psi_{M-1} + \psi_M) + \cdots + B_{-1}e^{r}(\psi_1 + \cdots + \psi_M)\right\} \end{aligned}$$

$$= \frac{1}{A_m} \sum_{j=0}^{M-1} e^{-rj} \sum_{x=1}^{M-j} B_{-x} p_{x+j} m_{x+j}. \tag{14}$$

The value B^* is directly determined by the initial birth sequence and the reproductive and survival age patterns.

6. Conclusion. Why, in plain words, does a population converge? The argument presented here is both simple and new. Once the population's tendency to grow is eliminated, by "dividing growth out" of the dynamics, the process of population replacement literally smoothes the generations out. Where childbearing, and thus the function ψ, is not concentrated at one age but is spread over several years, past humps and hollows in the birth sequence are thrown in together in the replacement process. They are averaged together—they smooth out.

Adding growth back means that a smooth exponential increase is reached in the long run—an exponential that is fully fixed given information from the initial birth sequence and the survival and reproductive patterns. And once the birth sequence reaches exponential increase, the age composition must assume its stable shape, no matter what it started as.

Acknowledgment. I should like to thank Joel Cohen, Griffith Feeney, John Hajnal, Nathan Keyfitz, Michael Stoto, and Ludo van der Heyden for valuable suggestions and comments on a previous draft.

Notes

1. For an alternative approach to smoothing or repeated averaging processes, based on renewal theory, see Feller [**3**, Vol. 1, Chap. 13].

2. To show this we must borrow some vocabulary from the elementary algebra of congruences. Let p be the smaller of p, q and choose K to be the integer $(p-1)q$. Now, given any k, choosing $b = 1, 2$ through $p-1$ in sequence will cause the expression $K + k - bq$ to run over the complete system of residues modulo p. If not, then for some different b_1, b_2, $K + k - b_1 q$ would be congruent to $K + k - b_2 q$ (modulo p), and by cancellation b_1 would be congruent to b_2 (modulo p), which is a contradiction. Therefore we can choose an integer b less than or equal to $p - 1$ so that $K + k - bq$ is nonnegative and has residue 0. Hence a nonnegative divisor, a, of $K + k - bq$ can be found. $\hat{B}_K$ therefore "reaches back" to all initial points and the theorem follows.

References

1. A. J. Coale, The Growth and Structure of Human Populations, Princeton University Press, Princeton, N.J., 1972.

2. J. E. Cohen, Ergodic theorems in demography, Bull. Amer. Math. Soc., 1 (1979) 275–295.

3. W. Feller, Introduction to Probability Theory and Its Applications, vol. 1, 3rd ed., Wiley, New York, 1968.

4. N. Keyfitz, Applied Mathematical Demography, Wiley, New York, 1977.

5. P. H. Leslie, On the use of matrices in certain population mathematics, Biometrica, 33 (1945) 183–212.

6. A. Lopez, Problems in Stable Population Theory, Office of Population Research, Princeton, N.J., 1961.

7. A. J. Lotka and F. R. Sharpe, A problem in age-distribution, Philosophical Magazine, Ser. 6, 21 (1911) 435–438.

8. A. J. Lotka, A contribution to the theory of self-renewing aggregates, with special reference to industrial replacement, Annals of Mathematical Statistics, 10 (1939) 1–25.

9. ———, The stability of the normal age distribution, Proc. of the National Academy of Sciences, 8 (1922) 339–345.

10. B. Parlett, Ergodic properties of populations I: The one-sex model, Theoretical Population Biology, 1 (1970) 191–207.

MISCELLANEA

61. Authorities are not yet quite agreed as to what a quaternion is or ought to be.—*Webster's Collegiate Dictionary*, Third Edition, 1926, s.v. "quaternion."

S. KOVALEVSKY: A MATHEMATICAL LESSON

KAREN D. RAPPAPORT
Mathematics Department, Essex County College, Newark, NJ 07102

Sofya Kovalevsky was a noted writer whose works include both fiction and nonfiction. She was also a political activist and a public advocate of feminism. In addition, she was a brilliant mathematician who made significant contributions despite the enormous educational and political obstacles that she had to overcome. Somehow her many achievements have been forgotten. In those few instances where her work has not been lost it has been denigrated by such studies as Felix Klein's history of nineteenth-century mathematics. Klein dismisses Kovalevsky's work in the following manner: "Her works are done in the style of Weierstrass and so one doesn't know how much of her own ideas are in them."[1] He finds something wrong with all her research and credits her with only one positive accomplishment, drawing Weierstrass out of his shell through their correspondence. It is time to set this record straight and to let the facts speak for themselves.

Sofya Krukovsky, known affectionately as Sonya, was born in Moscow in 1850. Her father, a Russian army officer, retired in 1858 and moved the family—Sofya, her older sister, Anyuta, and her younger brother, Fedya—to Palibino, an estate near the Lithuanian border.

After settling at Palibino, the household discovered that they had not brought a sufficient amount of wallpaper with them. Rather than travel a great distance to obtain new wallpaper, they decided to use old newspapers on the wall. Since only the nursery required the paper, this was deemed an adequate solution. However, while searching the attic for newspaper, they discovered paper of a better quality. On it were the lecture notes from a calculus course taken by General Krukovsky. This is how the nursery walls came to be covered with the calculus notes that, in her later years, Sofya claimed to have studied. Sofya often repeated this anecdote and enjoyed reporting how her calculus teacher exclaimed: "You have understood them as though you knew them in advance."[2]

Kovalevsky claimed that her interest in mathematics was aroused by her Uncle Peter, who would discuss numerous abstractions and mathematical concepts with her. When the family tutor, Joseph Malevich, read of this in Sofya's autobiographical work, *Memories of Childhood*, he was incensed. He wrote a long essay in a Russian newspaper explaining why *he* should receive credit for Kovalevsky's mathematical development. In response to this criticism Kovalevsky wrote the following tribute in "An Autobiographical Sketch": "It is to Joseph Malevich that I am indebted for my first systematic study of mathematics. It happened so long ago that I no longer remember his lessons at all.... It was arithmetic that Malevich taught best...I have to confess that arithmetic held little interest for me."[3]

Kovalevsky[4] studied mathematics against her father's wishes. When she was thirteen, she smuggled an algebra text into her room and studied it. When she was fourteen she taught herself trigonometry in order to study a physics book written by her neighbor Professor Tyrtov—trigonometry was necessary for the optics section, and the young Sofya taught herself without tutor or text. By constructing a chord on a circle, she was able to explain the sine function and to develop the other trigonometric formulas. When Professor Tyrtov saw her work, he was struck by its similarity to the actual mathematical development. Calling her a new Pascal, Tyrtov pleaded with the General to permit Sofya to study mathematics. After a year of exhortation, General Krukovsky relented and allowed Sofya to go to Petersburg to study calculus and other subjects.

After completing her studies in 1867, Sofya wanted to continue her education, but the Russian university system was closed to women. The only option for study was to go to Switzerland, but General Krukovsky would not allow his daughters to go abroad.

The author received her Ph.D. from New York University in 1975 under Professor H. Weitzner. She is a tenured faculty member at Essex County College in New Jersey. Her interests are in the history of women in mathematics, exploratory data analysis, computer applications in statistics, perturbation theory, and partial differential equations. —*Editors*

Sofya's older sister, Anyuta, felt imprisoned at Palibino and sought a way out. She found it through the radical politics of the times. This was a period of political ferment in Russia. The nihilists, feminists, and radicalists were all active, and their ideas were brought to Palibino by the local priest's son on his vacation from school in Petersburg. While they scandalized the neighborhood, these ideas had great influence on Anyuta, who in turn influenced Sofya. Anyuta joined a radical group that advocated higher education for women and promoted the concept of the "fictitious husband" to enable women to obtain more freedom. A married woman did not need her father's signature for a passport, and so a fictitious husband would enable Anyuta to travel abroad for her education. Anyuta and her friend Zhanna found a 26-year-old university student, Vladimir Kovalevsky, who agreed to marry one of them. Unfortunately, for Anyuta, she brought Sofya to one of their meetings. Vladimir became infatuated with Sofya and insisted on marrying her. After several secret meetings and much intrigue, General Krukovsky consented, and Vladimir and Sofya were married in September 1868.

Following their marriage, the Kovalevskys left for Petersburg to study, and to search for a husband for Anyuta. With little effort Sofya had won the freedom to pursue her education, the freedom and independence that Anyuta had been fighting so hard for. Sofya's feelings of guilt about this can be seen in her letters to Anyuta, who was still confined at Palibino. She wrote: "At times a strange anguish comes over me and I feel ashamed that everything is coming to me so easily and without any struggle."[5]

In Petersburg, Sofya received permission from the instructors to attend classes *unofficially*. She wrote to her sister: "...Lectures begin tomorrow and so my real life begins at 9 A.M. ...[Vladimir] and friends will solemnly escort me by way of the backstairs so that there is hope of hiding from the administration and from curious stares."[6] It was at Petersburg that Sofya decided to concentrate on mathematics. In a letter to Anyuta she said: "I have become convinced that one cannot learn everything and one life is barely sufficient to accomplish what I can in my chosen field."[7]

The Kovalevskys and Anyuta, who was still unmarried but chaperoned by the young couple, left for Europe in 1869. Sofya intended to study mathematics; and Vladimir, geology. Anyuta planned to pursue her revolutionary activities. Sofya and Vladimir settled in Heidelberg, but Sofya was not permitted to matriculate at the university. She appealed to both the faculty and the administration. A special committee was formed, and it was decided that each individual professor could choose whether to permit Sofya Kovalevsky to attend his lectures unofficially.

Kovalevsky was now able to attend lectures, and her outstanding mathematical ability became the talk of Heidelberg. As a firm believer in education for women, she used her reputation to assist other Russian women in their efforts to attend the university. One of these women was her friend Yulya Lermontov, who later became the first female chemist in Russia. For many years Bunsen had described Sofya as "a dangerous woman"[8] because, according to him, Sofya had tricked him into permitting Yulya to use the previously all-male chemistry labs. In 1874, Karl Weierstrass asked Sofya for confirmation of the story because "he [Bunsen] writes fiction even if he doesn't publish it."[9]

Sofya, Vladimir, and Yulya lived and studied together in Heidelberg until the fall of 1869. When Anyuta arrived for a visit, she was quite surprised to find Sofya still living with her "fictitious husband" and proceeded to evict Vladimir from the apartment. A short time later, Vladimir left Heidelberg to study at Jena.

As her mathematics progressed, Kovalevsky felt the need to study with Karl Weierstrass, the most noted mathematician of the time, at the University of Berlin. She traveled to Berlin for the start of the fall semester of 1870, only to find the university closed to women. Sofya wrote: "The capital of Prussia proved to be backward. Despite all my pleadings and efforts I had no success in obtaining permission to attend the University of Berlin."[10]

Determined to study mathematics, Kovalevsky personally presented herself to Weierstrass as an aspiring student. On the basis of recommendations from Koenigsberger at Heidelberg, Weierstrass was willing to see her. He assigned her a set of problems on hyperelliptic functions

that he had just given to his class. Weierstrass was so impressed with the ability she demonstrated in her solutions that he personally requested the university to allow her to attend his classes unofficially. However, the university was intransigent in its decision.

Not wanting to waste this mathematical talent, Weierstrass offered to tutor Kovalevsky privately. The lessons, begun in the fall of 1870, lasted for four years. The working relationship lasted a lifetime. When Weierstrass made his offer, he had no idea that Kovalevsky would become the closest of his disciples and remain so until her death.

The lessons began with twice-weekly meetings, Sundays at Weierstrass's home and weekdays at Kovalevsky's. The mathematics lessons are partially documented in a series of forty-one letters from Weierstrass to Kovalevsky, spanning the period of March 1871 to August 1874. They show that the emphasis was on Weierstrass's favorite topic: Abelian functions. Kovalevsky's responses are unavailable, because Weierstrass had burned them on hearing of her death. However, some idea of the importance of this correspondence to Kovalevsky can be found in her writings: "These studies had the deepest possible influence on my entire career in mathematics. They determined finally and irrevocably the direction I was to follow in my later scientific work: all my work has been done precisely in the spirit of Weierstrass."[11]

These letters also show the development of a close personal relationship between Kovalevsky and Weierstrass. They give a glimpse of the growing affection that Weierstrass felt for his pupil. Weierstrass was unaware of Kovalevsky's marriage arrangement and did not understand Vladimir's appearances. It was not until two years later (October 1872) that Weierstrass learned the truth. The series of letters at this time indicate that the topic of a relationship between Kovalevsky and Weierstrass was raised and the truth about her marital status was made known. Both were upset at the scene, and Weierstrass made several attempts to reassure his pupil that he would thereafter only discuss science. For a short period the correspondence remained strictly mathematical, but it did not stay that way. Weierstrass was not able to mask his concern and affection for Kovalevsky.

In October of 1872, Weierstrass had suggested several possible topics for Kovalevsky's dissertation. By 1874, she had completed three original works, any one of which Weierstrass felt would be acceptable. Now he needed to find a university that would award Kovalevsky a degree. In July 1874, the University of Göttingen awarded Sofya Kovalevsky a Ph.D., in absentia, summa cum laude, without either orals or defense. Needless to say, this was an unprecedented event.

The three papers presented for the degree were:

1. "On the Theory of Partial Differential Equations"
2. "On the Reduction of a Certain Class of Abelian Integrals of the Third Rank to Elliptic Integrals"
3. "Supplementary Remarks and Observations on Laplace's Research on the Form of Saturn's Rings"

The first of the papers, on partial differential equations, was published in Crelle's journal in 1875. This was considered a great honor, especially for a novice mathematician, since Crelle's journal was considered the most serious mathematical publication in Germany.

In the first paper, Kovalevsky had generalized a problem that had been posed by Cauchy. Cauchy had examined an existence theorem for partial differential equations, and Kovalevsky generalized Cauchy's results to systems of order r containing time derivatives of order r. The mathematician H. Poincaré said that "Kovalevsky significantly simplified the proof and gave the theorem its definitive form."[12] Today, this theorem on the existence and uniqueness of solutions of partial differential equations is often, but not always, known as the Cauchy-Kovalevsky Theorem. While studying the partial differential equation problem, Kovalevsky examined the heat equation. Some of her results were helpful to Weierstrass and he wrote: "So you see, dear Sonya, that your observation (which seemed so simple to you) on the distinctive property of partial differential equations...was for me the starting point for interesting and very elucidating researches."[13]

Shortly after Sofya received her degree from Göttingen, her friend Yulya was awarded a degree in chemistry from Göttingen. Vladimir had received his degree in paleontology two years earlier.

Although she had earned her Ph.D. and had written a highly acclaimed paper, Sofya Kovalevsky was unable to get a job. Even the efforts of her mentor, Weierstrass, were fruitless. So, in 1875, together with Vladimir, Anyuta, and Anyuta's husband, Sofya returned to the family home in Russia. Weierstrass encouraged her to relax at home and "enjoy the pleasures of big city life. I know you won't give up your scientific work."[14] However, Kovalevsky did not actively pursue her work that winter. She felt guilty and wrote: "I worked far less zealously than I had done in Germany and, indeed, the situation was far less propitious for scholarly work... The only thing that still gave me scholarly support was the exchange of letters and ideas with my beloved teacher Weierstrass."[15] As important as this exchange may have been, the correspondence with Weierstrass stopped in 1875. It was resumed for a short time in 1878 but did not become regular until 1880. Weierstrass was very hurt by this neglect. However, it must be acknowledged that Kovalevsky was never a good correspondent, even while in Germany, so that it is not surprising to find this lapse in writing after her return to Russia.

In Russia, Kovalevsky again tried to find a job. The only position in mathematics available to a woman was teaching arithmetic in the lower grades of a girls' school, and since Sofya admitted that she was "unfortunately weak in the multiplication table"[16] she could not seriously consider such a position. Therefore, Kovalevsky turned to other intellectual pursuits. She wrote fiction, theater reviews, and popular-science reports for a newspaper. She was instrumental in the organization of the Bestuzhev School for Women, but because of what were considered her radical views she was not permitted to teach there.

During this period in Russia, General Krukovsky died and left Sofya a small inheritance. Vladimir invested this money in business enterprises that eventually went bankrupt.

By 1878, Sofya had become bored with her activities and wanted to return to mathematics. She wrote to Weierstrass for advice. Weierstrass was excited by this letter, the first he had received from his pupil in three years. However, Kovalevsky's return to mathematics was delayed by the birth of a daughter, Sofya Vladimirovna, in October of 1878.

On their return to Russia the Kovalevskys had assumed the obligations of a real marriage. This was done partly as an obligation to Sofya's parents and partly because of their new politics. It was their feeling to end lying relationships of all kinds, and so the marriage was finally consummated.

Kovalevsky's return to mathematics was encouraged by a scientific conference held in Petersburg in 1880. The Russian mathematician Chebyshev invited Kovalevsky to present a paper at this conference. She found her unpublished dissertation on Abelian Integrals, translated it from German to Russian in one night, and presented it to the conference. Although it had lain untouched for six years, it was well received by the mathematicians.

Following her presentation, the Swedish mathematician Gösta Mittag-Leffler, who had met Kovalevsky earlier while she was a student in Germany, offered to find her a position in his country. Kovalevsky was very appreciative of this offer and wrote, in 1881, to Mittag-Leffler: "[If I can teach] I may in this way open the universities to women, which have hitherto only been open by special favor, a favor which can be denied at any moment."[17]

Kovalevsky's desire for a position was spurred not only by her feminism but also by her need to do mathematics. She wished "at the same time, to be able to live for my work, surrounded by those who are occupied with the same questions."[18]

While waiting for word from Sweden, Kovalevsky looked in Berlin for research work. While Vladimir was on a business trip, Sofya secretly visited Weierstrass. When she decided that she would return to Berlin to pursue research, Vladimir was quite angry with the decision and their marriage ended. For the next two years, Kovalevsky lived a student's life in Berlin, and the care of her daughter was shared by Sofya's friend Yulya and her brother-in-law Alexander.

While Sofya Kovalevsky was busy conducting her research on the refraction of light in crystals, Vladimir Kovalevsky again managed to get himself into financial difficulties. There was a stock scandal, and Vladimir, faced with ruin, committed suicide in the spring of 1883. Sofya took solace in her mathematics and hoped to find a badly needed job in Stockholm.

Mittag-Leffler had recently been appointed the head of the mathematics department at the

newly founded University of Stockholm and was able to offer Sofya a position there. However, the job had conditions. In order to prove her competence, Kovalevsky was to teach for a probationary year, with no pay and no official university affiliation. Kovalevsky agreed to this because she had no other options.

In the fall of 1883, Sofya Kovalevsky arrived in Stockholm to become a lecturer at the University of Stockholm. Her reception was mixed. Although hailed by many, others agreed with A. Strindberg, who wrote in the local paper, "A female professor is a pernicious and unpleasant phenomenon—even, one might say, a monstrosity."[19]

In the spring of 1884, Kovalevsky lectured in German on partial differential equations. These lectures were well received, and Mittag-Leffler was able to obtain the funds for her appointment as a Professor of Higher Analysis in July 1884. Word of the appointment was sent to Sofya in Moscow, where she was spending the summer with her daughter. However, Sofya did not bring her daughter to Stockholm in the fall because she was still unsure of her position. She was publicly criticized for her child-care arrangements but chose to ignore this. She didn't bring her daughter to Sweden until 1885, when the child was eight years old.

In addition to joining the Stockholm faculty in 1884, Kovalevsky became an editor for *Acta Mathematica* and published her first paper on crystals. In 1885, she received a second appointment to the Chair of Mechanics. She also published a second paper on the propagation of light in crystals but was embarrassed when Volterra found a serious error in her work. She had used a multi-valued function as if it were a single-valued one. Since this research was performed after a three-year hiatus in her mathematical career, Kovalevsky felt that her teacher, Weierstrass, should have caught the error prior to publication. Weierstrass, quite distraught, blamed illness and overwork. (It might be added that Weierstrass was 70 years old at the time.)

While in Stockholm, Sofya lived for a time in Mittag-Leffler's home, where she met and developed a friendship with his sister, Anna Leffler. Anna Leffler, a well-known advocate of women's rights and a writer, encouraged Sofya's literary leanings. In 1887 they collaborated on a play entitled *The Struggle for Happiness*. It was based on an idea that had occurred to Sofya while she sat at the bedside of her dying sister.

After Anyuta died in the fall of 1887, Sofya felt lonely and despondent. The sisters had been close, and Sofya felt the loss deeply. However, at this time two events occurred that helped to assuage her grief. Both the announcement of a new competition for the Prix Bordin and the arrival in Stockholm of a Russian lawyer named Maxim Kovalevsky were to have profound effects on the life of Sofya Kovalevsky.

Early in 1888 the French Academy of Science announced a new competition for the Prix Bordin. Papers on the theory of the rotation of a solid body would be considered for the prize competition at the end of the summer. Gösta Mittag-Leffler encouraged Sofya to work on a paper for the competition.

While Sofya was engaged in her research on the paper, Maxim Kovalevsky arrived to give a series of lectures at Stockholm University. He had been dismissed from Moscow University for criticizing Russian constitutional law. Aside from politics, Sofya and Maxim had many interests in common, and their attraction resulted in a scandalous affair. Eventually Maxim proposed marriage, on the condition that Sofya give up her research. Even if she had wished to give up her mathematics, Sofya was too far into her work for the prize competition to stop. In order to free Sofya to do her work, Mittag-Leffler invited Maxim to his summer home in Uppsala. This was a wise move, for, as Sofya stated, "If burly Maxim had stayed longer, I do not know how I should have got on with my work."[20]

With Maxim gone, Sofya was able to finish her work and the paper was submitted on time. Of the fifteen papers, which were submitted anonymously, one was considered so outstanding that the award was increased from 3,000 francs to 5,000 francs. The Prix Bordin was awarded to Sofya Kovalevsky in December of 1888. Sofya attended the awards ceremony with Maxim. Special recognition was given to her work. In his congratulatory speech, the President of the Academy of

Sciences said: "Our co-members have found that her work bears witness not only to profound and broad knowledge, but to a mind of great inventiveness."[21]

Prior to Sofya Kovalevsky's work the only solutions to the motion of a rigid body about a fixed point had been developed for the two cases where the body is symmetric. In the first case, developed by Euler, there are no external forces, and the center of mass is fixed within the body. This is the case that describes the motion of the earth. In the second case, derived by Lagrange, the fixed point and the center of gravity both lie on the axis of symmetry of the body. This case describes the motion of the top. Sofya Kovalevsky developed the first of the solvable special cases for an unsymmetrical top. In this case the center of mass is no longer on an axis in the body. She solved the problem by constructing coordinates explicitly as ultra-elliptic functions of time. Kovalevsky continued this work in two more papers on a rigid body motion. These both received awards from the Swedish Academy of Sciences in 1889. Her later works on the subject have been lost.

Kovalevsky's professorship in Sweden was due to expire in 1889. Desirous of returning to her native country, she inquired about a position in Russia. Her request was flatly denied. Russian mathematicians were indignant at this slight of Kovalevsky and decided to honor her. It was suggested that an honorary membership in the Russian Academy of Sciences would provide that recognition. However, in order to do that, the charter had to be amended to allow for female membership. In November 1889, Chebyshev sent the following telegram to Kovalevsky: "Our Academy of Sciences has just now elected you a corresponding member, having just permitted this innovation for which there has been no precedent until now. I am very happy to see this fulfillment of one of my most impassioned and justified desires."[22]

Kovalevsky sought positions throughout Europe but was again unsuccessful. She therefore had to accept the renewal of the professorship in Stockholm.

While working in Stockholm, Kovalevsky regularly commuted to France, where she visited Maxim at his villa. During these visits Maxim encouraged her literary interests. She wrote *Memories of Childhood* in Russian. It was translated into Swedish and published in 1889. In order not to shock Swedish society with its personal revelations, it was released as a novel entitled *The Raevsky Sisters*. The original was published in Russian in 1890. Kovalevsky's novel *A Nihilist Girl* was written in Swedish in 1890. A Russian version had been started, but Sofya's sudden death left it unfinished. Maxim edited the two versions and was responsible for the posthumous publication of the novel. This book was highly praised by critics in Russia and Scandinavia.

It was Kovalevsky's frequent trips to France to visit Maxim that eventually caused her death. On returning to Stockholm from a visit, early in 1891, she fell ill. The illness was misdiagnosed, and by the time it was finally found to be pneumonia it was out of control. On February 10, 1891, Sofya Kovalevsky died. Although there was widespread mourning and eulogies were given around the world, the Russian Minister of the Interior, I. N. Durnovo, was concerned that too much attention was being paid to "a woman who was, in the last analysis, a Nihilist."[23]

The mathematical world was more generous in its praise. Mittag-Leffler gave the official eulogy for the University of Stockholm. Speaking of her as a teacher he said: "We know with what inspiring zeal she explained [her] ideas... and how willingly she gave the riches of her knowledge."[24] In his eulogy, Kronecker, of the University of Berlin, spoke of Kovalevsky as "one of the rarest investigators."[25] Karl Weierstrass, who felt her loss most deeply, having burned all of her letters, said " 'People die, ideas endure': it would be enough for the eminent figure of Sofya to pass into posterity on the lone virtue of her mathematical and literary work."[26]

In her short lifetime, Sofya Kovalevsky left a notable collection of political, literary, and mathematical works. Her contributions, completed in spite of many obstacles, certainly warrant her a place in our intellectual and mathematical history.

During her mathematical career Sofya Kovalevsky published ten papers in mathematics and mathematical physics. Three of these papers,[27, 28, 29] were written during her student days under Weierstrass (1870–1874). The articles on the refraction of light[30, 31, 32] were written years later in

Berlin (1881–1883) after her return to her mathematical researches. The remaining research on rigid body motion and Bruns's Theorem[33-36] was completed during her tenure at the University of Stockholm (1883–1891). The only complete collection of Kovalevsky's works is published in Russian.[37] Portions of her work in partial differential equations and rigid body motion appear in English, and since these are Kovalevsky's most important works they will be discussed in some detail.

The proof of the first general existence theorem in partial differential equations was presented by Cauchy in 1842, in his publications on integrating differential equations with initial conditions. Here he showed the existence of analytic solutions for ordinary differential equations and certain linear partial differential equations.

The Cauchy problem for the ordinary differential equation $du/dt = f(t,u)$, with the initial conditions $u = u_0$ and $t = t_0$ states:

> If $f(t,u)$ is an analytic function in a neighborhood of the point (t,u) there exists a unique solution of $du/dt = f(t,u)$ with the given initial conditions.

For systems of first-order partial differential equations of the form

$$\frac{\partial \mu_i}{\partial t} = F_i\left(t, x_1, \ldots, x_n; u_1, \ldots, u_n; \frac{\partial u_1}{\partial x_1}, \ldots, \frac{\partial u_m}{\partial x_m}\right)$$

with initial conditions when $t = 0$, that is $u_i(0, x_1, \ldots, x_n) = w_i(x_1, \ldots, x_n)$ for $i = 1, \ldots, m$, the Cauchy problem is to find a solution $u(x, t)$ that satisfies the initial conditions.[38]

To solve this problem Cauchy assumed that F_i and w_i were analytic. He obtained a locally convergent power series solution by using his "method of majorants." The original function F_i is replaced by a simple analytic function whose power series expansion coefficients are nonnegative and greater than or equal to the absolute value of the corresponding coefficients for F_i. The derived system is then explicitly integrated to give a solution which is the majorant for the solution to the original with $t = 0$.

In her thesis Kovalevsky generalized Cauchy's results to systems of an order r containing time derivatives, $\partial^r u_i/\partial t^r$, of order r. It is this generalization that is known as the Cauchy-Kovalevsky Theorem.

Kovalevsky used majorants of the form

$$\frac{M}{1 - [(t_1 + t_2 + \cdots + t_r)/\rho]},$$

where ρ and M are constants.

For the higher order system

$$\frac{\partial^{n_i} u_i}{\partial t^{n_i}} = F_i\left(t, x_1, \ldots, x_n; u_n; u_1, \ldots, u_m; \frac{\partial^k u_j}{\partial t^{k_0} \partial x_1^{k_1} \ldots \partial x_n^{k_n}}\right)$$

with $i, j = 1, 2, \ldots, m$; $k_0 + k_1 + \cdots + k_n = k \leqslant n_j$, $k_0 < n_j$, and initial values given at some point $t = t_0$ for the u_i and their first $n_i - 1$ derivatives with respect to t, Kovalevsky proved the following.

If all the functions F_i are analytic in a neighborhood of the point $(t, x_1, \ldots, x_n; \ldots w_{j, k_0, k_1, \ldots, k_n})$ and all the functions $w_j^{(k)}$ are analytic in a neighborhood of the point $(x_1, \ldots, x_n)$, then Cauchy's problem has an analytic solution in a certain neighborhood of the point $(t, x_1, x_2, \ldots, x_n)$, and it is the unique solution in the class of analytic functions. (Note that

$$w_{j, k_0, k_1, \ldots, k_n} = \partial^{k-k_0} u_j / \partial x_1^{k_1} \ldots \partial x_n^{k_n}.)^{39}$$

The simplest form of this theorem states that any equation of the form $\partial u/\partial t = f(t, x, u, \partial u/\partial x)$,where f is analytic in its arguments for values near $(t_0, x_0, u_0, \partial u_0/\partial x_0)$, possesses one and only one solution $u(t, x)$ which is analytic near (t_0, x_0).

The Cauchy-Kovalevsky Theorem has significant limitations. It is restricted to analytic functions, and convergence may fail in a region of interest. Also the computation of the coefficients of the series may be too tedious to be practical. However, it is important in that it shows that within a class of analytic solutions of analytic equations the number of arbitrary functions needed for a general solution is the same as the order of the equation and these arbitrary functions involve one less independent variable than the number occurring in the equation.

The work for which Kovalevsky was best known in her time was her research on the motion of a rigid body about a fixed point. The equations of motion of a rigid body moving about a fixed point were derived by Euler in 1750. They are as follows:

$$\begin{aligned}
\frac{d\gamma}{dt} &= r\gamma' - q\gamma'' & \frac{Adp}{dt} + (C - B)qr &= Mg(y_0\gamma'' - z_0\gamma'),\\
\frac{d\gamma'}{dt} &= p\gamma'' - r\gamma & \frac{Bdq}{dt} + (A - C)rp &= Mg(z_0\gamma - x_0\gamma''),\\
\frac{d\gamma''}{dt} &= q\gamma - p\gamma' & \frac{Cdr}{dt} + (B - A)pq &= Mg(x_0\gamma' - y_0\gamma).
\end{aligned}$$

Here A, B, C are the principal axes of the ellipsoid of the body relative to the fixed point; M is the mass; g is the acceleration due to gravity. $\gamma, \gamma', \gamma''$ are the direction cosines of the angles which the three axes make at each moment; their direction is the same as the force of the rigid body; p, q, r are the components of the angular velocity along the principal axes; x_0, y_0, z_0 are the coordinates of the center of gravity of the body considered in a system of coordinates of which the origin is the fixed point and whose direction coincides with that of the principal axes of the ellipsoid of inertia.

The problem to be solved was the integration of this system of equations so that the position of the moving body at any time could be obtained. Before 1888, the integration had been completed for only two cases. The first was for the condition $x_0 = y_0 = z_0 = 0$. This case, studied by Euler and Poisson, is the one where the center of gravity of the moving body coincides with the fixed point. This is the motion of a force-free symmetric body. There are no external forces acting on the body and the motion is about a fixed point within the body, the center of mass. If the fixed point is the center of gravity, then gravity does not influence the motion. The axes of rotation are thus fixed in the body. An example of this force-free motion is the free rotation of the earth. In the case of the earth's free rotation, the axis of rotation does not coincide with the axis of symmetry. It is very slightly tilted, although it passes through the center of mass of the earth. What then happens is that the instantaneous axis precesses slowly about the axis of symmetry.

In the second case, studied by Lagrange, $A = B$, $x_0 = y_0 = 0$. Here the fixed point and the center of gravity lie on the same axis. When this axis is the axis of symmetry, the motion is that of the spinning top. "A top is defined to be a material body which is symmetrical about an axis and terminates in a sharp point ... at one of the axis."[40] This top spins about a fixed point that is not the center of gravity, but the center of gravity and the fixed point both lie on the axis of symmetry of the top. The weight of the top gives rise to a moment of force as it spins about the fixed point on the plane.

In both of these cases the rigid body was symmetrical. Sofya Kovalevsky developed the first of the soluble special cases for the heavy or unsymmetrical top. In this case, the center of mass no longer lies on the axis of the body. Instead, it is in the equatorial plane (the plane perpendicular to the axis) and passes through the fixed point. In addition, two of the principal moments of inertia are equal and double the third. The center of gravity is in the plane of the equal moments of inertia.

Euler's equations containing six unknown functions has the following first integrals:

$$Ap^2 + Bq^2 + Cr^2 - 2Mg(x_0\gamma + y_0\gamma' + z_0\gamma'') = C_1$$
$$Ap\gamma + Bq\gamma' + Cr\gamma'' = C_2$$
$$\gamma^2 + \gamma'^2 + \gamma''^2 = C_3 = 1.$$

It is sufficient to find one more integral to obtain a complete solution in quadratures. This occurs because the time variable appears only in the form dt and can be eliminated, leaving only five equations.

Kovalevsky derived the fourth integral for the case $A = B = 2C$, $z_0 = 0$,[41] and showed that the only conditions necessary for a given series to be a solution to Euler's system are that the constants A, B, C, x, y, z satisfy one of four conditions:

(1) $A = B = C$,
(2) $x_0 = y_0 = z_0 = 0$ (Euler's case),
(3) $A = B$, $x_0 = y_0 = 0$ (Lagrange's case),
(4) $A = B = 2C$, $z_0 = 0$ (Kovalevsky's case).[42]

Kovalevsky obtained a solution for her case by rotating the coordinate axes in the xy plane and choosing a unit of length so that $y_0 = 0$ and $C = 1$. With $c_0 = Mgx_0$ the Euler equations become:

$$2\frac{dp}{dt} = qr \qquad \frac{d\gamma}{dt} = r\gamma' - q\gamma''$$
$$2\frac{dg}{dt} = -pr - c_0\gamma'' \qquad \frac{d\gamma'}{dt} = p\gamma'' - r\gamma$$
$$2\frac{dr}{dt} = c_0\gamma' \qquad \frac{d\gamma''}{dt} = q\gamma - p\gamma'.$$

The three algebraic integrals are:

$$2(p^2 + q^2) + r^2 = 2c_0x + 6l_1$$
$$2(p\gamma + q\gamma') + r\gamma'' = 2l$$
$$\gamma^2 + \gamma'^2 + \gamma''^2 = 1$$

where l and l_1 are constants of integration. Kovalevsky then derived a fourth integral:

$$\left[(p + qi)^2 + c_0(\gamma + i\gamma')\right]\left[(p - qi)^2 + c_0(\gamma - \gamma' i)\right] = k^2$$

where k is an arbitrary constant. Defining $x_1 = p + qi$, $x_2 = p - qi$, she made several transformations of the variables. After some algebraic manipulations she obtained the equations:

$$0 = \frac{ds_1}{\sqrt{R_1(s_1)}} + \frac{ds_2}{\sqrt{R_1(s_2)}}$$
$$dt = \frac{s_1 ds_1}{\sqrt{R_1(s_1)}} + \frac{s_2 ds_2}{\sqrt{R_1(s_2)}},$$

where $R_1(S)$ is a fifth-degree polynomial whose zeros are unique and s_1 and s_2 are polynomials in x_1 and x_2. This system results in hyperelliptic integrals which Kovalevsky solved by using theta functions.

For this highly praised research Kovalevsky was awarded the Bordin Prize in 1888 and a prize from the Swedish Academy in 1889. Historians, however, have found this work "not of sufficient interest for the theory of a top to find a place here."[43] Today the value of her work is seen not in the results or the originality of the method but in the interest it stimulated in the problem of the rotation of a rigid body. Several researchers have continued the work of finding new cases of

special solutions. This includes Chaplygin's development of the integral for a symmetrical top turning about a fixed point, with moments of inertia $A = B = 4C$ and the Hesse-Schiff equations of motion for a top. There is also a body of work by others, for example, Bobylev, Steklov, Goryachev, and V. Kovalevsky.

The remaining works by Sofya Kovalevsky are of lesser importance and will be only briefly reviewed.

In the paper on Abelian integrals, Kovalevsky showed how a certain type of Abelian integral could be expressed as an elliptic integral.

The paper on Saturn's rings is concerned with the stability of motion of liquid ring-shaped bodies. Laplace found the form of the ring to be a skewed cross-section of an ellipse. Kovalevsky, using a series expansion, proved that the rings were egg-shaped ovals symmetric about a single axis. However, the subsequent proof that Saturn's rings consist of discrete particles and not a continuous liquid made this work inapplicable.

In the articles on the refraction of light in crystals, Kovalevsky applied a method developed by Weierstrass to differentiate Lamé's partial differential equations. However, Volterra discovered a basic error in her work. She had (as had Lamé) treated a multi-valued function as though it were single-valued, and the solution could not be applied to the equations in her form. However, the paper did demonstrate the previously unpublished theory of Weierstrass.

Kovalevsky's last article derived a simpler proof of Bruns's Theorem based on a property of the potential function of a homogeneous body.

Notes

1. Felix Klein, *Vorlesungen Uber die Entwicklung der Mathematik Im 19. Jahrhundert*, Teil I, Springer-Verlag, Berlin, 1926, p. 294.
2. Sofya Kovalevskaya, "An Autobiographical Sketch," in *A Russian Childhood*, transl. and ed. by Beatrice Stillman, Springer-Verlag, New York, 1978, p. 216.
3. *Ibid.*
4. Although Kovalevskaya, the feminine version of the name Kovalevsky, is used in much of the current literature, the author will refer to Kovalevsky in this paper, as that was the name under which Sofya Kovalevsky worked and published. However, in the notes the author has used the version of the name that appeared in the reference.
5. S. V. Kovalevskaya, *Vospominaniya i pis'ma*, ed. S. Ya. Shtraikh, Moscow, 1951, p. 232. Cited in Beatrice Stillman, "Introduction," in Kovalevskaya, *A Russian Childhood*, p. 11.
6. Kovalevskaya, *Vospominaniya i pis'ma*, p. 225. Cited in Stillman, "Introduction," in Kovalevskaya, *A Russian Childhood*, p. 11.
7. P. Polubarinova-Kochina, *Sophia Vasilyevna Kovalevskaya, Her Life and Work*, Men of Russian Science, transl. by P. Ludwick, Foreign Languages Press, Moscow, 1957, p. 25.
8. Karl Weierstrass, *Briefe von Karl Weierstrass an Sofie Kowalevskaja 1871–1891—Pis'ma Karla Veiershtrassa k Sof'ye Kovalevskoi*, Nauka, Moscow, 1973, p. 51.
9. *Ibid.*
10. Kovalevskaya, "An Autobiographical Sketch," in *A Russian Childhood*, p. 218.
11. *Ibid.*
12. P. Y. Polubarinova-Kochina, "On the Scientific Work of Sofya Kovalevskaya," transl. by N. Koblitz, in Kovalevskaya, *A Russian Childhood*, p. 234.
13. *Ibid.*, p. 235.
14. Weierstrass, *Briefe*, p. 52.
15. Kovalevskaya, "An Autobiographical Sketch," in *A Russian Childhood*, p. 220.
16. L. A. Vorontsova, *Sofia Kovalevskaia*, Moscow, 1957, p. 147. Cited in B. Stillman, "Sofya Kovalevskaya: Growing Up in the Sixties," *Russian Literature Triquarterly*, 9 (1974) 287.
17. Anna Carlotta Leffler, Duchess of Cajanello, *Sonya Kovalevsky*, transl. by A. de Furuhjelm, T. Fisher Unwin, London, 1895, p. 51
18. *Ibid.*, p. 52.
19. Polubarinova-Kochina, *Sophia V. Kovalevskaya*, p. 50.
20. S. Kovalevskaya to A. Leffler, 1888. Cited in Anna Leffler, *Sonya Kovalevsky*, p. 124.
21. Polubarinova-Kochina, *Sophia V. Kovalevskaya*, p. 61.

22. Kovalevskaya, *Vospominaniya i pis'ma*, p. 352. Cited in G. J. Tee, "Sof'ya Vasil'yevna Kovalevskaya," *Mathematical Chronicle*, 5 (1977) 132–133.
23. Kovalevskaya, *Vospominaniya i pis'ma*, p. 532. Cited in Tee, p. 135.
24. G. Mittag-Leffler, Commemorative Speech as Rector of Stockholm University, 1891. Cited in Leffler, p. 171.
25. L. Kronecker, *Journ. für die reine und angewandte Mathematik*, BD CVIII, March 1891, 1–18.
26. G. Mittag-Leffler, "Weierstrass et Sonja Kowalewsky," *Acta Mathematica*, 39 (1923) 170.
27. S. Kovalevsky, "Zur Theorie der partiellen Differentialgleichungen," *Journal für die reine und angewandte Mathematik* 80 (1875) 1–32.
28. ______, "Über die Reduction einer bestimmten Klasse von Abel'scher Integrale 3-ten Ranges auf elliptische Integrale," *Acta Mathematica*, 4 (1884) 393–414.
29. ______, "Züsatze und Bermerkungen zu Laplace's Untersuchung uber die Gestalt des Saturnringes," *Astronomische Nachrichten* 111 (1885) 37–48.
30. ______, "Über die Brechung des Lichtes in cristallinischen Mitteln," *Acta Mathematica*, 6 (1883) 249–304.
31. ______, "Sur la propagation de la lumière dans un milieu cristallisé," *Comptes rendus des séances de l'académie des sciences*, 98 (1884) 356–357.
32. ______, "Om ljusets fortplantning uti ett kristalliniskt medium," *Öfversigt af Kongl. Vetenskaps-Akademiens Forhandlinger*, 41 (1884) 119–121.
33. ______, "Sur le problème de la rotation d'un corps solide autour d'un point fixe," *Acta Mathematica*, 12 (1889) 177–232.
34. ______, "Mémoire sur un cas particulier du problème de la rotation d'un corps pesant autour d'un point fixe, où l'intégration s'effectue à l'aide de fonctions ultraelliptiques du temps," *Mémoires présentés par divers savants à l'académie des sciences de l'institut national de France*, 31 (1890) 1–62.
35. ______, "Sur une propriété du système d'équations différentielles qui définit la rotation d'un corps solide autour d'un point fixe," *Acta Mathematica*, 14 (1890) 81–93.
36. ______, "Sur un théorème de M. Bruns," *Acta Mathematica*, 15 (1891) 45–52.
37. S. V. Kovalevskaya, *Nauchnye raboty* (Scientific Works), ed. by P. Y. Polubarinova-Kochina, AN SSSR, Moscow, 1948.
38. G. Birkhoff, ed., *A Source Book in Classical Analysis*, Harvard University Press, Cambridge, Mass., 1973, p. 318.
39. P. Y. Polubarinova-Kochina, "On the Scientific Work of Sofya Kovalevskaya," transl. by N. Koblitz, in Kovalevskaya, *A Russian Childhood*, p. 233.
40. E. T. Whittaker, *Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, London, 1965, p. 164.
41. Kovalevsky, "Mémoire sur un cas particulier du problème de la rotation d'un corps pesant..." (see Note 34).
42. Kovalevsky, "Sur une propriété du système..." (see Note 35).
43. A. Gray, *A Treatise on Gyrostatics and Rotational Motion, Theory and Applications*, Dover Publications, New York, 1959 (first appeared 1918), p. 369.

THE UNIFORMIZATION THEOREM

WILLIAM ABIKOFF
Department of Mathematics, University of Illinois, Urbana, IL 61801

Almost one hundred years have passed since Felix Klein discovered the uniformization theorem. While such theorems had earlier been proved in specific cases, no one had dared even conjecture that every compact Riemann surface could be parametrized by a variable whose

William Abikoff received his B.S. and M.S. in electrical engineering and his Ph.D. in mathematics from the Polytechnic Institute of Brooklyn. He worked at Bell Telephone Laboratories and taught at Columbia University and the universities of Perugia, Firenze, and Paris. He is currently an associate professor at the University of Illinois at Urbana-Champaign. He has been a research fellow at the Mittag-Leffler Institute and the Institut des Hautes Études Scientifiques. His research interests have centered on the use of geometric function theory in the study of Riemann surfaces, Teichmüller theory, and Kleinian groups. Currently he is studying the structure at infinity of hyperbolic manifolds.—*Editors*

domain of definition lay in the Riemann sphere, $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. For reasons that will become apparent in § 3, Klein asserted much more, calling his result the limit circle theorem. The theorem occurred to him at 2:30 A.M. on March 23, 1882, while he was in the midst of an asthma attack. His health had been poor and he was trying to recover in the North Sea coastal town of Nordeney. The combination of miserable weather and a desire to announce the theorem led him to return to his native city of Düsseldorf almost immediately. When he received the galley proofs of the announcement from the publisher, he dispatched copies to Hurwitz, Schwarz, and Poincaré; Hurwitz was also sent an outline of the proof. The responses accorded this result by his distinguished colleagues are classic examples of the camaraderie of scientists—at least in my loose interpretation of Klein's recollections [**14**, p. 584].

HURWITZ: I accept it without reservation.
SCHWARZ: It's false.
POINCARÉ: It's true. I knew it and I have a better way of looking at the problem.

However, the story does not end there. Schwarz recanted shortly thereafter. He made two great contributions to the theory of uniformization. The first was a method for proving the theorem. The second was to establish the relationship between uniformization and the study of conformal representation of Riemann surfaces on $\hat{\mathbb{C}}$. The link is provided by the theory of covering spaces which Schwarz developed specifically to study the uniformization problem. In 1907, Poincaré proved the most general known uniformization theorem for the case where the parameter varies over a simply connected domain.

The result, which is now called the uniformization theorem, is a problem in conformal mapping which is solved by potential theory. It is related to the original problem by algebraic topology. Of the major mathematical disciplines, few have not been enriched by the uniformization theorem or the methods developed for the study of the problem.

Research in uniformization theory has gone through several dormant periods since the concept of a (global) uniformization was introduced by Klein in 1882. Uniformization theorems of power unimaginable to the classical masters have been proved in the past twenty years and no end is in sight.

My original purpose in writing this paper was to put into print a relatively efficient and elementary proof of half of the classical uniformization theorem which I have circulated privately for several years. Ralph Boas suggested that I write it in a more expository form. I have taken this suggestion as an invitation to exercise my personal fascination with the various facets of the uniformization problem. To do historical justice to the problem, I was required to trace the development of the notion of a Riemann surface; for as this notion matured, so did the statements and proofs involved in the solution of the uniformization problem. The discussion in the text contains an outline of this proof with the details placed in an appendix. For completeness I have also included a brief sketch of Maskit's work on the general uniformization problem.

In translating heuristic arguments into mathematics, I shall often refer the reader to Ahlfors's *Conformal Invariants* [**2**], which will be abbreviated as CI.

I would like to express my thanks to Lars Ahlfors, Lipman Bers, Jozef Dodziuk, and Bernard Maskit for their comments, many of which are incorporated in the ensuing text. The figures, which, in a sense, are the soul of Riemann surface theory, were designed and drawn by George Francis. My debt to him is clear.

1. An Example. Consider the variety S in $\mathbb{C}^2$ defined by the equation $X^2 + Y^2 - 1 = 0$. Here X and Y are complex and S is the solution set. There is an obvious method for parametrizing S. For $z \in \mathbb{C}$, set $X(z) = \cos z$ and $Y(z) = \sin z$. S is completely parametrized or, in the language we will use here, S is uniformized by the variable z.

This is usually considered a poor choice of uniformization for the following reasons. $X(z)$ and $Y(z)$ are transcendental functions with essential singularities at ∞. The structure at ∞ is not

clearly displayed by the uniformization.

Another choice is to let z vary over $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$

$$\text{with } X(z) = 2z/(1+z^2) \quad \text{and} \quad Y(z) = (1-z^2)/(1+z^2).$$

The behavior at ∞ is not exceptional; however, the rational uniformization has poles at $\pm i$, i.e., there is a definite structure at ∞. If we adjoin to S the ideal points $(X(\pm i), Y(\pm i))$, we obtain a compact Riemann surface $\bar{S}$. $\bar{S}$ is homeomorphic to $\hat{\mathbb{C}}$ as may be seen from Riemann's method of branch cuts.

To eliminate the special role of the point at ∞, the variety S may be viewed as lying in the complex projective plane by homogenizing the original polynomial, i.e., we consider the variety in projective space defined by $X^2 + Y^2 - W^2 = 0$ (see, for example, Kendig [**12**]).

2. The Evolution of the Concept of a Riemann Surface. Riemann surfaces were first introduced by Riemann in an attempt to understand multivalued functions of complex variables. The equation $X^2 - Y^2 = 0$ does not define X as a single valued function of Y. However, if X and Y are chosen to have real values, a choice of sign of $\sqrt{Y}$ may be consistently made if Y is not permitted to assume the value zero. At $X = Y = 0$, the differential of $P(X, Y) = X^2 - Y^2$ vanishes, that is, $(0,0)$ is a singularity of the equation $P(X, Y) = 0$. Avoiding the singularities is not sufficient to permit such choice if X and Y are permitted to have complex values.

Using a construction that may be found in any elementary complex variables text, Riemann resolved the problem by constructing sheets above the complex plane. On each sheet, a choice may be made. The sheets are then glued together to form a natural domain S_1 of definition for the function $\sqrt{X^2}$. In Fig. 1, we demonstrate Riemann's solution for real values of X and Y; Fig. 2 gives the local picture in $\mathbb{C}$ near 0. Notice that the modification is made on the domain of the function not on the range. Riemann initially thought of his sheets as lying over the complex plane in Euclidean 3-space. Much of the subsequent development of the notion of Riemann surface was done to remove the artificiality and arbitrary character of Riemann's embedding in $\mathbb{R}^3$. As Weyl noted, S_1 is a natural parameter space for the variety S. In this sense Riemann had a uniformization theorem, but it was not the first.

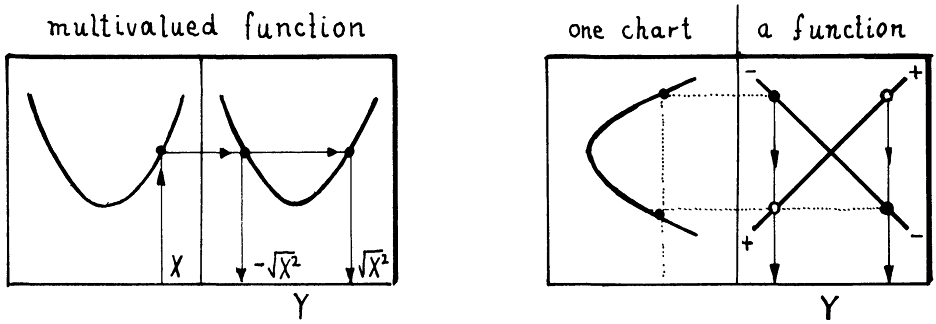


FIG. 1

As far as I know, the first uniformization theorem was proved by Puiseux in 1850. The theorem is a local uniformization theorem for the structure of singularities of plane algebraic curves. It states that if X and Y satisfy a polynomial relation $P(X, Y) = 0$, then, up to a linear change of coordinates, X can be written locally as a holomorphic function of $y = (Y - Y_0)^{1/k}$ for some $k \geqslant 1$; y then becomes a local uniformizing parameter for the variety $P(X, Y) = 0$. A proof may be found in Hille [**11**, vol. 1, p. 265 ff.].

Algebraic equations in two variables are not the only source of multivalued functions. Linear differential equations with regular singular points, such as $XY' + Y = 0$, generally have multivalued solutions.

Algebraic integrals are contour integrals in the complex plane of the form $\int R(z, w)\, dz$ where R

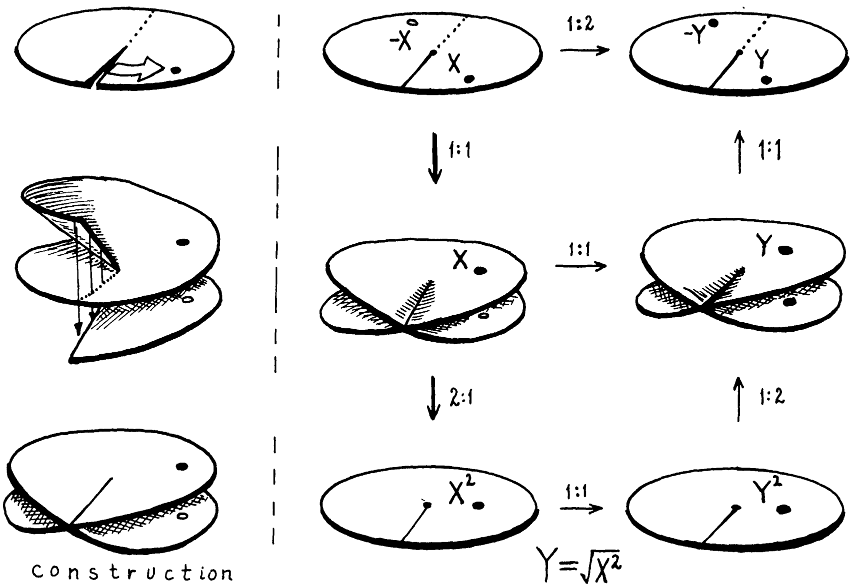


FIG. 2

is rational and z and w satisfy a polynomial identity. As the path of integration varies, we again obtain multivalued functions of the path and upper endpoint of integration. Siegel [**21**, pp. 1 ff.] traces back to Fagnano in 1719 the study of the algebraic properties of these integrals. In the first half of the nineteenth century, these properties were studied by Abel and Jacobi; the integrals are now called abelian integrals.

Riemann later studied these integrals by considering them as defined on the variety S rather than as integrals in the complex plane.

The method discovered by Weierstrass for studying multivalued functions is quite different from Riemann's. It is probably the first step in the development of the modern notions of abstract manifold and sheaf. Weierstrass's original paper on the subject seems to have been written in 1842, although it did not appear in print until 1894. The basic idea is to construct a Riemann surface, in this context called an *analytic configuration*, by starting with a power series $f(z)$ centered at some z_0 of positive radius of convergence. One then considers all meromorphic continuations of f in $\hat{\mathbb{C}}$. Each continuation of f to a point z is given by a chain of discs in which the continuation is possible.

The set of points to which f may be continued is subject to an equivalence relation, and a natural topology may be given to the equivalence classes (see, e.g., Ahlfors [**1**] or Hille [**11**, vol. 2] for details). Singular points admitting uniformizations by Puiseux series are usually added to give *complete analytic configurations*, which are Riemann surfaces in the sense of Weyl (see below). The analytic configurations are Riemann surfaces which lie in the sheaf of germs of holomorphic or meromorphic functions on $\hat{\mathbb{C}}$. Overlapping discs define the topology, a concept which underlies the modern notion of manifold.

In his 1907 paper on uniformization, Poincaré removed the reliance of Weierstrass's construction on the initial power series $f(z)$. He proved the uniformization theorem for collections of overlapping discs on which there exists *some* meromorphic function in the sense of Weierstrass. The geometric entity became paramount; the role of the function was secondary.

The modern era in the study of Riemann surfaces, and, indeed, of all manifolds, opened in 1913 with the publication of Hermann Weyl's book *The Concept of a Riemann Surface*. Here, for the first time, no defining function is assumed. The definition is purely geometric and also independent of any embedding in Euclidean 3-space.

On the justification for abandoning Riemann's embedding of the Riemann surfaces of multivalued functions, Weyl wrote:

> In essence, three-dimensional space has nothing to do with analytic forms [Riemann surfaces of multi-valued functions], and one appeals to it not on logical-mathematical grounds, but because it is closely associated with our sense perception. To satisfy our desire for pictures and analogies in this fashion by forcing inessential representations on objects instead of taking them as they are could be called an anthropomorphism contrary to scientific principles. However, these reproaches of the pure logician are no longer pertinent if we pursue the other approach ... in which the analytic form is a two-dimensional manifold ... To the contrary, not to use this approach is to overlook one of the most essential aspects of the topic.

Weyl's definition, in modern language and without unnecessary hypotheses, follows. Let S be a connected Hausdorff space and $U = \{D_\alpha | D_\alpha \subset S\}$ be an open cover of S. Assume that for all α there is a homeomorphism $z_\alpha^{-1}: D_\alpha \to \mathbb{C}$ which satisfies: $z_\beta^{-1} \cdot z_\alpha$ is holomorphic where defined. The set $A = \{(z_\alpha, D_\alpha)\}$ is called a *holomorphic* or *conformal atlas* and the pair (S, A) is called a *Riemann surface*. Usually, by abuse of language, we speak of the Riemann surface S. z_α is called a *local coordinate* or *uniformizing* variable at the points of D_α. A function f on S is *holomorphic, meromorphic* or *harmonic* if, for each α, $f \circ z_\alpha$ has that property wherever it is defined. A map $f: S_1 \to S_2$ between two surfaces S_1 and S_2 is *holomorphic* if $z_2^{-1} \cdot f \cdot z_1$ is holomorphic whenever defined. Here z_1 and z_2 are local coordinates on S_1 and S_2, respectively.

It is not trivial, but possible, to prove that every Riemann surface in the sense of Weyl carries a pair of nonconstant linearly independent meromorphic functions and may be embedded in $\mathbb{R}^3$. It then follows that, up to holomorphic equivalence, the notions of Riemann surfaces developed by Riemann (suitably generalized), Weierstrass, and Weyl are equivalent.

To return to the uniformization problem, the question changed as the notion of Riemann surfaces changed. Klein's original claim is to have proved the theorem for compact Riemann surfaces, possibly missing a finite number of points, in the sense of Riemann. All proofs until, I believe, the 1920's assumed that the surfaces were first countable; this latter assumption was made by Weyl. The assumption was removed by Radó, but now may be derived as a simple consequence of the uniformization theorem.

We may now state the general uniformization problem. Let S be a Riemann surface. Find *all* domains $D \subset \hat{\mathbb{C}}$ and holomorphic functions $t: D \to S$ so that at each point $p \in S$, t is a local uniformizing variable at p.

Equivalently, there is a topological disc $B \subset S$ with center p so that the restriction of t to each component of $t^{-1}(B)$ is a homeomorphism. The reader should notice that the required conditions on t and D say precisely that the triple (D, S, t) is a (smooth) covering space with base S, total space D, and holomorphic projection t.

The early problem was less ambitious. It was simply to find one uniformization, but where D is simply connected. This theorem was proved independently by Koebe and Poincaré in 1907. Poincaré's solution was somewhat more general but we will ignore that generalization here. Until recently the most common proof was that offered by Hilbert in 1909. We give the modern statement of the Uniformization Theorem in the next section following the preliminaries that tie it to the uniformization problem.

3. Covering Surfaces and Classical Plane Geometries. Let S_1 and S_2 be two Riemann surfaces and $\pi: S_1 \to S_2$ be a local homeomorphism so that each point on S_2 has an evenly covered neighborhood. We then say that (S_1, S_2, π) is a *covering surface* or, more commonly, that S_1 is a *covering surface* of S_2 with *cover map* or *projection* π. The number of points in $\pi^{-1}(p)$ is independent of $p \in S_2$ and is called the *number of sheets* of the covering.

Covering surfaces occur classically in the study of algebraic equations with symmetries. For example, the equation $w^2 - z = 0$ admits the symmetry obtained by replacing w by $-w$. In this context the symmetry is called *sheet interchange*. The Riemann surface S of $w^2 - z = 0$ over $\mathbb{C} \setminus \{0\}$ is a covering surface of $\mathbb{C} \setminus \{0\}$. Notice that the sheet interchange cannot be defined to

be a covering surface at $z = 0$. Points where the map topologically looks like $z \mapsto z^n$ are called *ramification points*.

Covering surfaces with isolated singularities, all of which are ramification points, are called *ramified covers*.

A method for constructing a covering surface is to take two copies S' and S'' of a given surface S, slice them along corresponding curves, and glue as indicated in Fig. 3. The resulting surface is a two-sheeted cover of S with the projection taking z' and z'' to z. Notice that there is a simple closed curve that covers β twice. In a sense that may be made quite precise, passing to the covering surface has replaced β by a curve of twice the length of β. If we repeat this process infinitely often, β will have been replaced by an infinitely long, simple curve; i.e., β will no longer be an obstruction to simple connectivity. We may repeat this process for sufficiently many curves so that the resulting covering surface $\tilde{S}$ has only homotopically trivial curves. $\tilde{S}$ is called the *universal covering surface*. This approach is the most classical one; the modern approach (see, e.g., Greenberg [**8**]) is an abstract reformulation of the basic idea of opening up all homotopically nontrivial closed curves. Klein states that the construction above is due to Schwarz.

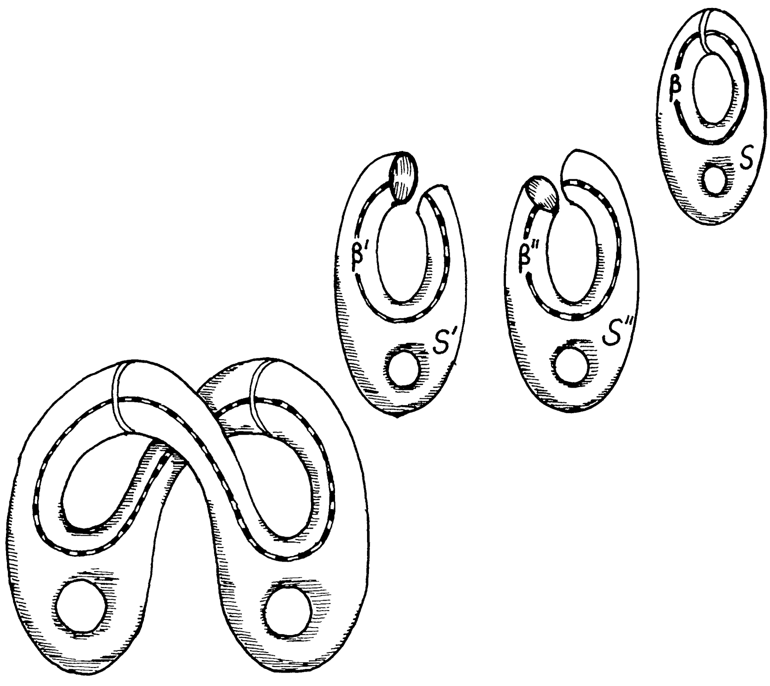


FIG. 3

As we have noted in the previous section, if S is the Riemann surface of some multivalued function Y of X, then both X and Y are functions of ζ if ζ varies over the points of S. If S_1 is a covering surface of S, and ζ_1 varies over S_1, then $X \cdot \pi$ and $Y \cdot \pi$ parametrize the variety in $\mathbb{C}^2$ defined by the functional relationship between X and Y. Thus if S_1 lies in $\hat{\mathbb{C}}$ we obtain a solution to the general uniformization problem. We will return to this question in § 6 but now restrict our attention to $S_1 = \tilde{S}$.

$\tilde{S}$ inherits the structure of a Riemann surface from the conformal structure on S. $\tilde{S}$ is a simply connected Riemann surface because any homotopically nontrivial closed curve has been opened. It follows immediately that we may obtain a solution to the uniformization problem by showing that every simply connected Riemann surface is biholomorphically equivalent to a subdomain of

$\hat{\mathbb{C}}$; this is merely a rephrasing of the Riemann mapping theorem, (see § 4). So a solution to the uniformization problem follows from the statement now known as:

THE UNIFORMIZATION THEOREM. *Every simply connected Riemann surface is biholomorphically equivalent to either* $\mathbb{C}$, $\hat{\mathbb{C}}$, *or the unit disc* Δ.

That the three possibilities are distinct is an immediate consequence of Liouville's theorem.

The universal cover has several very strong properties. Since it is simply connected, every multivalued locally meromorphic function on S lifts to a meromorphic function on $\tilde{S}$; here we are restating the monodromy theorem. The universal cover is the only covering surface with this property.

Another important property is the following. Suppose $\zeta \in S$ and N is an evenly covered neighborhood of ζ. The cover map $\pi: \tilde{S} \to S$ defines a map γ from one component N_1 of $\pi^{-1}(N)$ to another, say N_2, by the rule $\zeta_1 \mapsto \zeta_2$ where ζ_2 is the unique point in N_2 for which $\pi(\zeta_1) = \pi(\zeta_2)$ (see Fig. 4). Using the evenly covered character of N_1, one shows that γ extends to a biholomorphic self-map of $\tilde{S}$. These maps form a group G called the *group of the covering* $(\tilde{S}, S, \pi)$ or *cover group* or *group of deck* (*cover*) *transformations*. Clearly, G acts without fixed points. G also has the stronger property of being *properly discontinuous*, which means that each point $\tilde{\zeta} \in \tilde{S}$ has a neighborhood $\tilde{N}$ so that $\gamma(\tilde{N}) \cap \tilde{N} \neq \varnothing$ if and only if $\gamma =$ identity. Since the sheet interchange maps $\gamma \in G$ are in one-to-one correspondence with the free homotopy classes of closed curves on S, the group G is isomorphic to the fundamental group $\pi_1 S$ of S based at any point. The orbit space $\tilde{S}/G$ is the set of equivalence classes in $\tilde{S}$ where $\zeta_1 \sim \zeta_2$ if there exists $\gamma \in G$ so that $\zeta_2 = \gamma(\zeta_1)$. It is not difficult to show that $\tilde{S}/G$ inherits a conformal structure from $\tilde{S}$ and that S is biholomorphically equivalent to $\tilde{S}/G$.

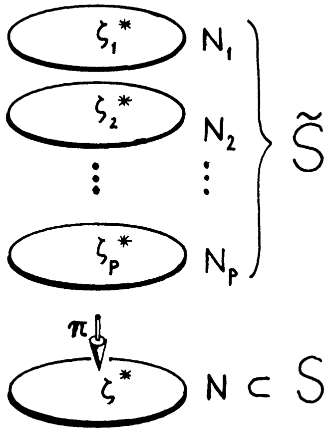


FIG. 4

The line of reasoning developed above is, more or less, historically accurate when $\tilde{S}$ is $\hat{\mathbb{C}}$ or $\mathbb{C}$ (we shall no longer distinguish between biholomorphically equivalent surfaces except when necessary). It shows immediately that $\tilde{S} = \hat{\mathbb{C}}$ only if $S = \hat{\mathbb{C}}$. If S is compact and $\tilde{S} = \mathbb{C}$, π is the inverse to an elliptic integral (see Siegel [**21**, Chapter 1]). The problem of finding inverse functions to algebraic integrals, i.e., cover maps, is called the Jacobi inversion problem, the rich history of which leads directly to Riemann surfaces and uniformization (cf. Weyl [**23**, p. 144]).

When $\tilde{S} = \mathbb{C}$, G must be a properly discontinuous subgroup of the group $\operatorname{Aut}\mathbb{C}$ of biholomorphic self-maps of $\mathbb{C}$. There are precisely three types. The first is when G is trivial. The Riemann surface is then $\mathbb{C}$. The second are the cyclic groups $\{z \mapsto z + nz_0 \mid n \in \mathbb{Z}\}$. All are conjugate in

$\operatorname{Aut}\mathbb{C}$ to $G = \{z \mapsto z + n \mid n \in \mathbb{Z}\}$ and conjugate groups determine biholomorphically equivalent surfaces. The surface is $\mathbb{C} \setminus \{0\}$ and $\pi : \mathbb{C} \to S$ is the map $\exp(2\pi iz)$. The last is a class of groups, the lattices, which are, up to conjugation, generated by $\gamma_1 : z \mapsto z + 1$ and $\gamma_2 : z \mapsto z + \tau$ where $\operatorname{Im} \tau > 0$. In this case $\pi_1 S$ is free abelian on two generators and S is topologically a torus. The two generators may be chosen as in Fig. 5.

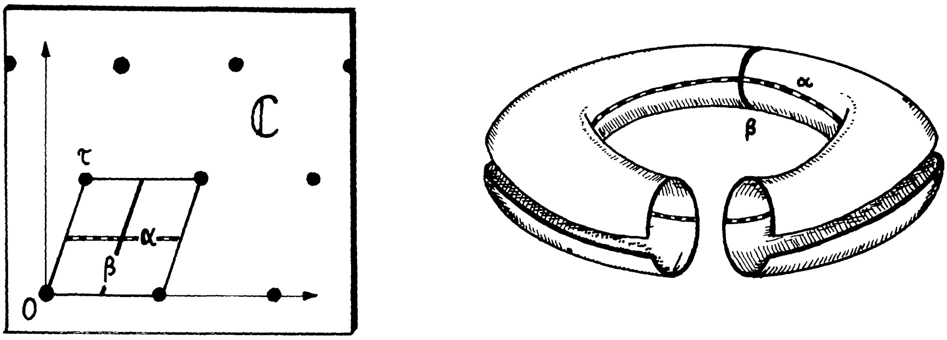


FIG. 5

We have this "found" all Riemann surfaces with $\tilde{S} = \mathbb{C}$ or $\tilde{S} = \hat{\mathbb{C}}$. The uniformization theorem tells us that every other Riemann surface S has $\tilde{S} = \Delta$. Here the history proceeded in the opposite direction. Klein knew in 1882 that if G is a properly discontinuous subgroup of $\operatorname{Aut}\Delta$ then Δ/G looks like a Riemann surface. However in 1882, a Riemann surface required a defining multivalued function. While it seems that Klein had a method for obtaining such a function, Poincaré claimed two methods, one of which remains the standard method for obtaining automorphic functions for discrete groups acting on bounded domains. The existence of this function shows that all the definitions of Riemann surfaces coincide.

The Riemann sphere has a metric of constant positive curvature induced by the usual embedding in $\mathbb{R}^3$. $\mathbb{C}$, hence by projection $\mathbb{C} \setminus \{0\}$ and tori, have Euclidean or flat metrics, i.e., of zero curvature. Using the uniformization theorem, we see that all other Riemann surfaces inherit any $\operatorname{Aut}\Delta$ invariant metric from Δ. Up to scale factor, there is only one; it is the Poincaré metric on the hyperbolic plane. It is a metric of constant negative curvature. The universal cover of a surface of constant negative curvature may be conformally and injectively developed on a hyperbolic plane. This is one of many successful approaches to proving the uniformization theorem. The relevant differential equation is $\Delta u = e^{2u}$. Studies of this equation were among the earliest in nonlinear partial differential equations.

We also recall that the hyperbolic plane, in Poincaré's model, is the unit disc Δ. When we map $\tilde{S}$ to Δ, we give natural boundary to $\tilde{S}$, namely $\partial\Delta$. To Klein the edge was a limiting circle and he called his uniformization theorem, the *limit circle theorem*.

To comment on the pre-1907 "proofs" of the uniformization is to tread on the most dangerous ground. One gathers from Hilbert's 1900 lecture at the International Congress [**10**] that he did not accept Klein's proof, for it is not mentioned. Poincaré's uniformization is spoken of, but Hilbert comments that it does not parametrize the whole variety; so it is not a uniformization in the sense considered here (see also the introduction to Poincaré [**20**]).

From the vantage point of 1980, it is not quite so easy to dismiss Klein's argument. One is first presented with a major obstacle, namely, to find the argument. To those of us trained in the Satz-Beweis school of mathematical exposition and discourse, reading Klein is often a mystical experience. In fact, many a mathematician has proved and published a deep and elegant result, later to discover with chagrin that there is a casual and vague reference to the result in Klein. Most of Klein's writing on areas related to uniformization are collected in two books written jointly with Fricke; these comprise over 2,000 pages without an index, and often without

definitions or theorem statements. The mathematical insight contained therein is astounding, but it often seems that one can only appreciate a part of it after having independently rediscovered the results. To the modern observer Klein only claimed the proof of the uniformization theorem when S is *conformally finite*, that is, when S is a compact surface missing a finite number of points. A terse modern appraisal of the argument is that it is an excellent outline, but far from a proof; it appears on pages 698–705 of volume 3 of his collected works [**14**].

A brief description follows. First construct one algebraic equation defining $P(z, w) = 0$ as a multivalued function of z. Let S be the Riemann surface of P in the sense of Riemann. On S we find a finite set of piecewise circular closed curves α_i so that $S \setminus \cup \alpha_i$ is a polygon $\Pi \subset S$. On S, z is a well-defined function, $z|\Pi$ immerses Π in $\hat{\mathbb{C}}$, and $\partial z(\Pi)$ consists of circular arcs. P may be chosen in any genus g, so that it is very symmetric. Then for some choice of P and the α_i, $z(\Pi) \subset \Delta$ and $\partial z(\Pi)$ are circular arcs lying on circles orthogonal to $\partial\Delta$. $z^{-1}|z(\Pi)$ may be analytically, but in a multivalued fashion, continued along all paths in S. If the image domain is the unit disc, the continuation is the inverse of the universal cover map. He then considers the corresponding group of cover transformations, and he notes that the space of such (normalized) groups G and the space of dissected Riemann surfaces in genus g both have real dimension $6g - 6$. The local correspondence between them he *assumes* is a local real analytic diffeomorphism. He, more or less, shows that the mapping is injective and proper, hence bijective. This completes Klein's attempt. This technique of proof is called the *continuity method*. Even assuming the uniformization theorem, the last two properties are true but not easily proved. The fundamental difficulty with this proof was recognized quite early and goes as follows. Forget that the correspondence is a local diffeomorphism (or wait some 40 to 80 years until the theorem is proved). You only have a proper, continuous injection f of $\mathbb{R}^n$ into $\mathbb{R}^n$. Brouwer essentially developed dimension theory to prove that f is a homeomorphism and thereby resurrected the continuity method. With Brouwer's proof added, Klein's technique becomes viable; however, before Brouwer's proof appeared, the uniformization theorem had already been proved in complete generality.

Uniformization theory was relatively dormant from 1883 to 1900. In 1900, Hilbert delivered a lecture to the International Congress of Mathematicians in which he stated 23 problems which have had a profound effect on the course of mathematics in the twentieth century. Uniformization was Problem 22. This renewed interest in the question led to the solution in 1907. Both solutions given in 1907 and the argument given in 1909 by Hilbert come from potential theory and it is to that stream of ideas that we now turn.

4. Some Potential Theory. Riemann's thesis is contemporary with many of the great discoveries of nineteenth-century physical science. Klein [**13**, p. x] wrote, "Riemann as we know used Dirichlet's Principle in their place." The physical arguments of which Klein speaks are those associated to a conservative vector field E and its associated potential function V. The classic examples of these fields are electric fields and the flow of an incompressible fluid. I shall not dwell on these concepts save to say that the following equations hold: $\operatorname{div} E = 0$, $E = -\operatorname{grad} V$, and $\Delta V = 0$. In the last equation, $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ is the Laplacian, and therefore V is a harmonic function.

Among the earliest nontrivial fields to be studied is that induced by a point charge. In the unit disc, the potential defined by a (positive) point charge at $z = 0$ is $V(z) = \log(1/|z|)$ if $\partial\Delta$ is grounded, i.e., $V|\partial\Delta = 0$. Riemann considered the question of whether a point charge could live in an arbitrary simply connected plane domain D whose boundary is a grounded conductor.

The latter condition is that the potential $V|\partial D = 0$. Assume we have a point charge in D. The level curves $L_V(c)$ of V for $0 < c < \infty$ are analytic Jordan curves separating the point charge at z_0 from ∂D. The integral curves C of the gradient field of V, the paths of elections in this field, are again analytic curves from ∂D to z_0. These integral curves may be parametrized by the angle θ at which they enter z_0. We may therefore write $C = C(\theta)$. Each point $z \in D$ lies on a unique level curve $L_V(c)$ and a unique integral curve $C(\theta)$. We form a map

$$f: D \to \Delta$$
$$z \mapsto (e^{-c}, \theta)$$

in polar coordinates. (See Fig. 6.) This map is conformal and proves the Riemann mapping theorem, once we know that a point charge can live in D. To prove existence, Riemann invoked the Dirichlet principle. The principle states that harmonic functions minimize the energy in an electric field and such a minimum exists here if $D \neq \mathbb{C}$. This argument was used by Riemann in several contexts, but was questioned by Weierstrass. The latter noted that even elementary extremum problems need have no solution. The Dirichlet principle fell into disrepute but was later resurrected by Hilbert. It is the key to his 1909 proof of the uniformization theory. Hilbert's proof uses mapping properties associated with electric dipoles.

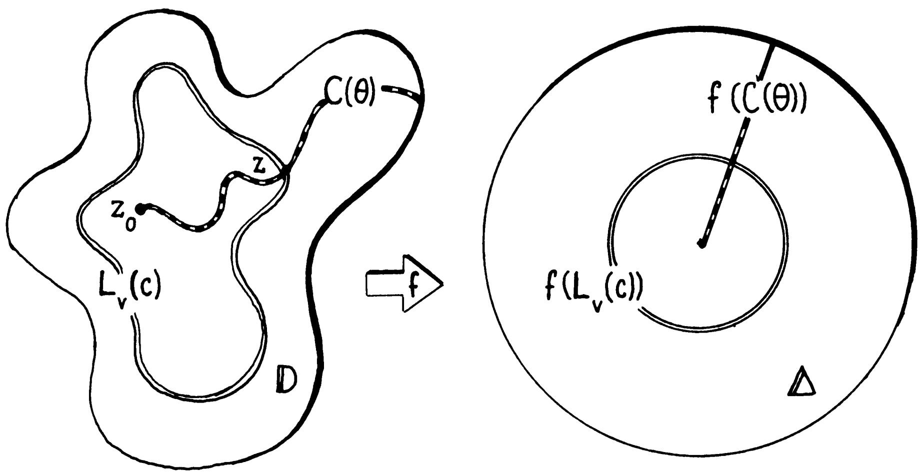


FIG. 6

To continue our intuitive discussion, we next consider incompressible fluid flow. (See Fig. 7.) If we have a point source of water in the plane, for example a faucet, there must be an edge across which the water may flow out of the domain. Otherwise the fluid will compress. If the domain has such an edge, which we call a "thick boundary," then a point source can exist at any point in the

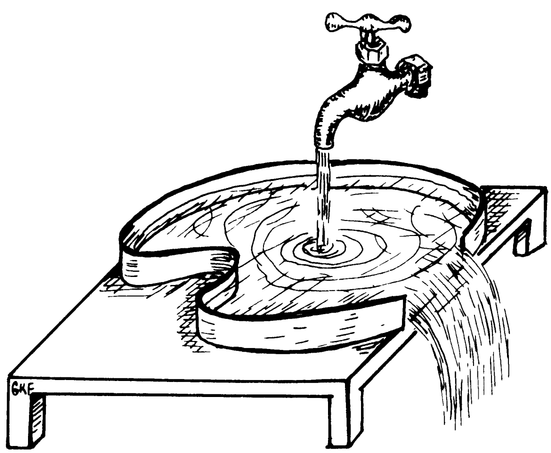


FIG. 7

domain. It is intuitively obvious that any edge save a point for the domain will enable it to support a point source.

There are now many ways to prove the existence of a point charge or source. Perhaps the most elegant is due to Perron. His method may be found in most complex analysis books. Basically the idea is that a harmonic function is the 2 (or n) variable analogue of a linear function $f(x)$ of one variable. $f(x)$ has the property that $f(x) = \sup g(x)$ where $g(x)$ is convex and the boundary values of g are less than or equal to those of f (see Fig. 8). Locally one replaces g by a linear function and g is "bootstrapped" up to f by taking suprema.

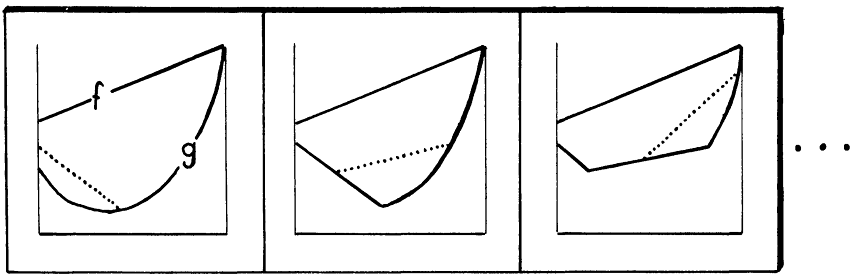


FIG. 8

On a Riemann surface, subharmonic functions assume the role of convex functions. A subharmonic function is a continuous real-valued function whose value at a point $\zeta_0 \in S$ is less than or equal to the average of the values of the function on the boundary of any small symmetric neighborhood of ζ_0. Specifically, let $s\colon S \to \mathbb{R}$ be continuous. s is *subharmonic* if, for each $\zeta_0 \in S$ and each small neighborhood N and local coordinate $z = z(\zeta)$ in N with $z(\zeta_0) = 0$,

$$s(\zeta_0) \leqslant \frac{1}{2\pi}\int_0^{2\pi} s\left(re^{i\theta}\right)d\theta$$

for $r = |z(\zeta)|$ sufficiently small and $\theta = \arg z(\zeta)$.

A *Perron class* $\mathfrak{F}$ is a nonvoid set of subharmonic functions $s\colon S \to \mathbb{R}$ which is closed under the following operations:

(i) taking the maximum of two functions

(ii) local harmonic majorization.

Property (ii) generalizes the local replacement of a convex function by a linear function. To obtain the precise definition, let $\bar{D}$ be a closed disc on S. On $\bar{D}$, let $\hat{s}$ be the harmonic function such that $\hat{s}|\partial D = s|\partial D$. Extend $\hat{s}$ to S by setting $\hat{s} = s$ in $S \setminus D$. $\hat{s}$ is called the *local harmonic majorant* of s in D. *Local harmonic majorization* is the process of replacing s by $\hat{s}$ for some disc D.

Perron showed that $\bar{s}(\zeta) = \sup_{s\in\mathfrak{F}} s(\zeta)$ is either harmonic or identically infinite.

We now give a precise definition of the Green's function, or potential of a point charge, using Perron's method. The potential should be the smallest potential which is positive and grows as does $-\log|z|$ near the point charge.

Let $\zeta_0 \in S$ and $\mathfrak{F}$ be the set of subharmonic functions s in $S \setminus \{\zeta_0\}$ having the following properties:

(i) $s(\zeta) \geqslant 0$.

(ii) $\{\zeta \in S \setminus \{\zeta_0\} | s(\zeta) \neq 0\}$ is compact.

(iii) For any local coordinate z on a neighborhood N of ζ_0 with $z(\zeta_0) = 0$, $s(z) + \log|z|$ is bounded above near ζ_0.

As a consequence of Perron's Theorem (see, e.g., Ahlfors [**1**, p. 240]), $\bar{s}(\zeta) = \sup_{s\in\mathfrak{F}} s(\zeta)$ is either identically infinite or harmonic in $S \setminus \{\zeta_0\}$. In the latter case we call $g(\zeta,\zeta_0) = \bar{s}(\zeta)$ the *Green's function* of S with singularity at ζ_0. In the former case we say that S does not admit a Green's function with singularity at ζ_0. It is known that, if z is a local coordinate at ζ_0 with

$z(\zeta_0) = 0$, $g(\zeta, \zeta_0) + \log|z|$ has a harmonic extension to any small neighborhood of ζ_0 (see CI, p. 141).

Assume that S has a Green's function $g(\zeta, \zeta_0)$. Then, in the sense of fluid flow, it must have a thick boundary—a notion we must also formalize. Asserting that a domain S has a thick boundary is equivalent to stating that the "edge" of S may serve as a source or sink for incompressible fluid flow (with bounded potential). Equivalently the associated potential may have a minimum along the edge. The nonthickness of the boundary is formalized by the following:

DEFINITION. Let S be a Riemann surface and K be a compact subset of S. We say that S satisfies the *maximum principle relative to* K if every bounded harmonic function f: $S \setminus K \to \mathbb{R}$ has the property that

$$\sup_{\zeta \in S \setminus K} f(\zeta) = \overline{\lim_{\zeta \to \partial K}} f(\zeta).$$

Notice that if S has a Green's function with singularity $\zeta_0 \in \operatorname{Int} K$, then $-g|(S \setminus K)$ shows that S does not satisfy the maximum principle relative to K. The converse is also true, namely, the invalidity of the maximum principle relative to K implies the existence of the Green's function with singularity at an arbitrary point $\zeta_0 \in \operatorname{Int} K$ (see CI, p. 139).

Given a Riemann surface or any plane domain, the validity of the relative maximum principle is not directly verifiable. There is, however, a standard technique used to show that it is not valid. Suppose there exists a nonconstant harmonic function h on S. Then h has a maximum on K. If the relative maximum principle were valid, that maximum would be a global maximum contradicting the usual maximum principle. We have proved:

PROPOSITION 4.1. *Let S be a Riemann surface. If there is a bounded harmonic function f: $S \to \mathbb{R}$ which is not constant, then, for all compact $K \subset S$ with nonvoid interior, the maximum principle relative to K is not valid and S has a Green's function with singularity at any point $\zeta_0 \in S$.*

It is not a triviality to produce nonconstant bounded harmonic functions on Riemann surfaces. For example, the usual maximum principle implies that a harmonic function on a compact Riemann surface is constant. $\mathbb{C}$ admits the nonconstant harmonic function $\operatorname{Re} z$, but, by applying the removable singularity theorem to a neighborhood of infinity, $\mathbb{C}$ admits no bounded nonconstant harmonic functions.

Producing nonconstant bounded harmonic functions on a Riemann surface brings us to the venerable Dirichlet problem. The problem is easily stated. Let $S \subset S_1$ be Riemann surfaces and, for simplicity, assume ∂S is a finite collection of piecewise analytic curves. Let $f: \partial S \to \mathbb{R}$ be continuous. The *Dirichlet problem* is to find a continuous function $h: \bar{S} \to \mathbb{R}$ so that $h|S$ is harmonic and $h|\partial S = f$. Notice that ∂S corresponds to our intuitive picture of a thick edge. Further, if we may solve the Dirichlet problem for nonconstant functions f, then, by the proposition above, S will not satisfy the maximum principle relative to a compact subset and our analytic characterization of thickness will have been proved.

The solution of the Dirichlet problem is an application of Perron's method. Assume f is bounded or, more simply, assume ∂S is compact. We form a Perron class $\mathfrak{F}$ of continuous functions $s: \bar{S} \to \mathbb{R}$ which are subharmonic in S and satisfy $s|\partial S \leqslant f$ and $s \leqslant \sup f$. If $m = \inf f$, then the constant function $m \in \mathfrak{F}$. Also, the maximum principle, applied to subharmonic functions, implies that for all $\zeta \in S$ and all $s \in \mathfrak{F}$, $s(\zeta) \leqslant \sup f$. It follows that $h(\zeta) = \sup_{s \in \mathfrak{F}} s(\zeta) < \infty$ and hence is harmonic in S. It remains to show that h extends to ∂S and $h|\partial S = f$. The usual technique for doing so is due to Poincaré (1899) and formalizes the notion of local thickness of the boundary. One aims a microscope with arbitrarily fine resolution at a point $\zeta \in \partial S$, and we see whether it is possible to push water across ∂S near ζ. The potentials of these flows are called *barriers*. The formal definitions and proofs may be found in Bers [4, p. 139 ff.] and Conway [7, p. 265 ff.]. Conway's arguments are stated for plane domains but are equally valid on Riemann surfaces. The precise result that we need is

PROPOSITION 4.2. *Let S_1 and S be Riemann surfaces and $S \subset S_1$. If ∂S is a finite union of closed analytic arcs, then, for all continuous, bounded $f: \partial S \to \mathbb{R}$ there exists $h: \bar{S} \to \mathbb{R}$ so that*

(i) $|h| \leq \sup|f|$

(ii) $h|S$ *is harmonic*

(iii) $h|\partial S = f$.

As an immediate consequence of Propositions 4.1 and 4.2, we obtain the first conclusion of

PROPOSITION 4.3. *Let S_1 and S be Riemann surfaces and $S \subset S_1$. If ∂S is a finite union of closed analytic arcs then*

(i) *S has a Green's function $g(\zeta, \zeta_0)$*

and

(ii) $\lim_{\zeta \to \partial S} g(\zeta, \zeta_0) = 0$.

Proof: We prove the second conclusion. Let C be a small circle around ζ_0. Using Proposition 4.2, on S we may solve the Dirichlet problem outside C with boundary data $f|C = g$ and $f|\partial S = 0$. Call the resulting solution h. By the maximum principle for subharmonic functions, h is an upper bound for all functions s lying in the Perron class $\mathfrak{F}$ defining $g(\zeta, \zeta_0)$. It follows that

$$\overline{\lim_{\zeta \to \partial S}}\, g(\zeta, \zeta_0) \leq 0.$$

Since the function $s(\zeta) \equiv 0$ lies in $\mathfrak{F}$,

$$\underline{\lim_{\zeta \to \partial S}}\, g(\zeta, \zeta_0) \geq 0,$$

which proves the Proposition.

5. The Uniformization Theorem. The uniformization theorem, even today, commands a nontrivial proof. Here we will sketch one style of proof with some details omitted. For the interested reader, we give references either to the accessible literature or to the appendix to this paper.

Let S be a simply connected Riemann surface. We first assume that, for fixed $\zeta_0 \in S$, S admits a point charge or Green's function $g = g(\zeta, \zeta_0)$ with singularity at ζ_0. The complete argument in this case may be found in CI, p. 136 ff. We must define a conformal map $f: S \to \mathbb{C}$. Let $S' = S \setminus \{\zeta_0\}$ and choose a simply connected chart N_ζ in S' near each $\zeta \in S'$. Since g is harmonic in N_ζ, it has a harmonic conjugate h_ζ there and $f_\zeta = \exp[-(g + ih_\zeta)]$ is holomorphic near ζ. Further, f_ζ is unique up to multiplication by a complex number of modulus one. Let N_0 be a simply connected chart near ζ_0 with local coordinate z satisfying $z(\zeta_0) = 0$. $g(\zeta, \zeta_0) + \log|z(\zeta)|$ is harmonic in N_0 and hence has a harmonic conjugate h_0. Set

$$f_0(\zeta) = z(\zeta) \cdot \exp\left[-\left(g + \log|z| + ih_0\right)\right].$$

f_0 is holomorphic in N_0 and vanishes to first order at ζ_0. By adjusting constants of modulus one, f_ζ is an analytic continuation of f_0. Since S is simply connected, the monodromy theorem implies that the analytic continuation defines a homomorphic function f: $S \to \mathbb{C}$. Also $|f(\zeta)| = e^{-g} < 1$ since $g(\zeta, \zeta_0) > 0$. It is possible to show directly that f is a bijection of S with Δ; however, an elegant argument due to Heins [**9**] is far more efficient. We omit the details save to note that the proof makes decisive use of the fact that a Riemann surface admitting a Green's function with singularity at ζ_0 also has a Green's function with singularity at any prescribed point. This completes our sketch of the proof of

PROPOSITION 5.1. *If S is a simply connected Riemann surface which admits a Green's function, then there is a biholomorphic, i.e., conformal, map f: $S \to \Delta$.*

Henceforth we assume that S is simply connected and does not admit a Green's function.

DEFINITION. A *divergent curve* on S is a piecewise analytic simple arc $\phi: [0, \infty) \to S$ so that,

for any compact $K \subset S$, $\phi^{-1}(K)$ is compact.

Now assume S admits a divergent curve and set $S_t = S \setminus \phi([t,\infty))$. It should be intuitively clear, but requires proof, that the simple connectivity of S implies that S_t shares that property. Using Proposition 4.3, we then obtain

LEMMA 5.1. *For all $t \geq 0$, S_t is simply connected and for any $\zeta_0 \in S_0$, S_t admits a Green's function with singularity at ζ_0. Further,*

$$\lim_{\zeta \to \partial S_t} g(\zeta,\zeta_0) = 0.$$

Proof: See the appendix.

We shall need the following standard result in function theory.

LEMMA 5.2. *Let $\Delta(r) = \{|z| < r\}$ and $\mathfrak{S}_r$ be the set of holomorphic injections $f\colon \Delta(r) \to \mathbb{C}$ and satisfying*
(i) $f(0) = 0$
(ii) $f'(0) = 1$.
Then $\mathfrak{S}_r$ is (sequentially) compact in the topology of uniform convergence on compact subsets.

Proof. The map

$$\mathrm{UC}\colon \mathfrak{S}_r \to \mathfrak{S}_1$$
$$f(z) \mapsto F(z) = r^{-1}f(rz)$$

is obviously a homeomorphism. It therefore suffices to show that $\mathfrak{S}_1$ is compact. Montel's Theorem (see any graduate-level complex analysis text) states that one must show only that $\mathfrak{S}_1$ is closed and bounded. Hurwitz's Theorem states that a limit of holomorphic injections is holomorphic and injective or constant. Condition (ii) rules out a constant limit. Thus $\mathfrak{S}_1$ is closed. The estimates necessary to show that $\mathfrak{S}_1$ is bounded are given by Koebe's distortion theorem (see CI, p. 84, or Conway [**7**, p. 351 ff.].

Since S_t is a simply connected Riemann surface with a Green's function, there is a holomorphic bijection $f_t\colon S_t \to \Delta$. Further we may assume that $f_t(\zeta_0) = 0$ for some fixed $\zeta_0 \in S$. Now choose a sequence (t_i) increasing to infinity and denote S_{t_i} by S_i and f_{t_i} by f_i. Fix the local coordinate $z = f_0(\zeta)$ near ζ_0. We may then compute $c_i = f_i'(z(\zeta))|_{\zeta=\zeta_0}$ and let $F_i(\zeta) = c_i^{-1}f_i(\zeta)$. $F_i\colon S_i \to \Delta_i = \Delta(c_i^{-1})$ is a holomorphic bijection.

Recursively we define subsequences N_i of $\mathbb{Z}^+$ as follows:
(i) $N_1 = \mathbb{Z}^+$
(ii) If N_i is defined and $j \in N_i$ and $j \geq i$, F_j is defined and injective on S_i and $F_j'(z(\zeta_0)) = 1$. $F_j \circ F_i^{-1}\colon \Delta_i \to \mathbb{C}$ is injective, maps 0 to 0, and has derivative equal to 1. Thus, by Lemma 5.2, we find that there must be a subsequence $N_{i+1} \subset N_i$ so that, for $j \in N_{i+1}$, $F_j \circ F_i^{-1}$ converges to an injective map $H_i\colon \Delta_i \to \mathbb{C}$. $H_i(0) = 0$ and $H_i'(0) = 1$. We have defined N_{i+1}.

Choose n_j to be the jth entry in the sequence N_j. For $k > i$, on S_i, $H_k \circ F_k$ is a holomorphic injection and

$$H_k \circ F_k = \left(\lim_{j\to\infty} F_{n_j} \circ F_k^{-1}\right) \circ F_k = \left(\lim_{j\to\infty} F_{n_j} \circ F_i^{-1}\right) \circ F_i \circ F_k^{-1} \circ F_k$$

$$= \left(\lim_{j\to\infty} F_{n_j} \circ F_k^{-1}\right) \circ F_i = H_i \circ F_i.$$

Thus $H_i F_i$ is the restriction to S_i of a globally defined holomorphic map $f: S \to \mathbb{C}$. f is injective since $f | S_i$ is injective for all i.

$f(S)$ is simply connected. If $f(S) \neq \mathbb{C}$, then, by the Riemann mapping theorem, $f(S)$ is conformally equivalent to Δ and there is a conformal map $h : S \to \Delta$. $\operatorname{Re} h$ is a bounded nonconstant harmonic function on S. As in Proposition 4.1, S must then have a Green's function which contradicts our original assumption. We have therefore proved

PROPOSITION 5.2. *If S is a simply connected Riemann surface with a divergent curve and admitting no Green's function, then S is conformally equivalent to $\mathbb{C}$.*

We shall need the following

PROPOSITION 5.3. *If S is a simply connected Riemann surface with no divergent curves, then, for all $\zeta_1 \in S$, $\dot{S} = S \setminus \{\zeta_1\}$ is simply connected.*

Proof. Here the reader is offered a choice of two proofs. A proof via potential theory and covering spaces is given in the appendix. The deepest but quickest proof uses the classification of simply connected topological surfaces. A simply connected Riemann surface S is homeomorphic to $\mathbb{C}$ or to $\hat{\mathbb{C}}$ (see Ahlfors and Sario [**3**, pp. 90–104]). In $\mathbb{C}$ it is easy to find a divergent curve ϕ. The image of ϕ in S may be arbitrarily closely approximated by a divergent curve. Otherwise S is homeomorphic to $\hat{\mathbb{C}}$, $\dot{S}$ is homeomorphic to $\mathbb{C}$ and hence is simply connected.

THE UNIFORMIZATION THEOREM. *If S is a simply connected Riemann surface, then S is conformally equivalent to Δ, $\mathbb{C}$ or $\hat{\mathbb{C}}$.*

Proof. If S has a Green's function then Proposition 5.1 shows that S is equivalent to Δ. If S has no Green's function but has a divergent curve, then S is equivalent to $\mathbb{C}$. In any other case, $\dot{S} = S \setminus \{\zeta_0\}$ is simply connected and obviously has a divergent curve. It follows that $\dot{S} \simeq \Delta$ or $\dot{S} \simeq \mathbb{C}$.

If $f: \dot{S} \to \Delta$ is a conformal equivalence, then f is a bounded holomorpic function on $\dot{S}$, in particular it is bounded near ζ_0. By Riemann's theorem on removable singularities, f extends to a holomorphic map of S into $\bar{\Delta}$. By the maximum principle, $|f(\zeta_0)| < 1$. Since $f(\dot{S}) = \Delta$, there is some $\zeta_1 \in \dot{S}$ so that $f(\zeta_0) = f(\zeta_1)$. By the open mapping theorem, there are points ζ_0', ζ_1' near ζ_0 and ζ_1, respectively, so that $f(\zeta_0') = f(\zeta_1')$. But this contradicts the fact that $f | \dot{S}$ is injective. Therefore $\dot{S} \simeq \mathbb{C}$ and $S \simeq \hat{\mathbb{C}}$ which completes the proof.

To illustrate the use of the uniformization theorem, we note

COROLLARY 1. *Every Riemann surface is second countable and separable.*

Proof. These properties project from the universal cover. Δ, $\mathbb{C}$, and $\hat{\mathbb{C}}$ have these properties.

COROLLARY 2 (Picard's Theorem). *Let f be a meromorphic function in $\mathbb{C}$. If $\hat{\mathbb{C}} \setminus f(\mathbb{C})$ contains at least three points, then f is constant.*

Proof. Let $D = f(\mathbb{C})$. By the uniformization theorem, the universal cover $\tilde{D}$ of D is conformally equivalent to Δ. Let π denote the universal cover map, $\pi : \Delta \to D$. Then, locally, π^{-1} exists and $\pi^{-1} \cdot f$ may be continued along all paths in $\mathbb{C}$ to define a map $F: \mathbb{C} \to \Delta$. Since F is holomorphic, Liouville's theorem says F is constant; hence so is f.

6. Maskit's Work on the General Uniformization Problem. In a fifteen-year series of papers, Maskit resolved the general uniformization problem for surfaces of finite conformal type. Weyl noted that the problem has two aspects. One starts with a Riemann surface S. The first part of the problem is topological—namely, find all covering surfaces $D \subset \hat{\mathbb{C}}$ of S. Maskit's planarity theorem [**12**] classifies them in the following way. On S we find a set of simple closed loops $\{\alpha_i\}$ which are homotopically independent. By this we mean that the α_i are disjoint and not freely

homotopic to each other or to the ideal boundary of S. To each α_t we assign a positive integer n_t. The set of pairs $P = \{(\alpha_t, n_t)\}$ determine a planar covering surface S_P which is defined as the largest covering surface on which $\alpha_t^{n_t}$ is a simple loop but α_t^k is not for $k < n_t$. This theorem gives a complete solution to the topological part of the problem.

The second part of the problem is the conformal mapping problem. Here we try to find all conformal maps of all surfaces S_p into $\hat{\mathbf{C}}$. The existence of such a map was proved by Koebe; it is his planarity theorem (see, for example, Tsuji [**18**]). Let f be the map and $D = f(S_p)$. The group of deck transformations G for the covering $D \to S$ lies in the conformal automorphism group Aut D of D. For purposes of classification, it is not important to know all conformal maps of S_p into $\hat{\mathbf{C}}$ but just one good one. All others are obtained by conformal maps of domains D in $\hat{\mathbf{C}}$. Maskit [**13**] found a very good one. Specifically he proved that D may be chosen so that for all $\gamma \in$ Aut D, γ is a Möbius transformation. The group G then becomes a group of Möbius transformations acting properly discontinuously on a domain $D \subset \hat{\mathbf{C}}$. Such objects had first been studied by Schottky and later by Fricke and Klein. They are called function groups. Now assume S has finite conformal type. Maskit further showed [**14**] D may be chosen so that each component of Int$(\hat{\mathbf{C}} \setminus D)$ is a Euclidean disc. Such groups, he called *Koebe groups*. In [**15**], he classified the Koebe groups. This solves the general uniformization for surfaces of finite conformal type.

Appendix

This appendix contains the proofs of Lemma 5.1 and Proposition 5.3. The proof of the former is rather short and we give it first.

Proof of Lemma 5.1. $S_t \subset S$ and ∂S_t is a piecewise analytic arc. As we noted in § 5, Proposition 4.3 implies that S_t has a Green's function for all $t \geqslant 0$. We claim that S_{t_0} is simply connected for all t_0. Let α be a closed curve in S_{t_0} based at ζ_0. Let

$$A(\alpha) = \{t \in [t_0, \infty) | \alpha \text{ is null homotopic in } S_t\}.$$

Since α is null homotopic in S, and the homotopy takes place in a compact subset of S, $A(\alpha) \neq \varnothing$ and is open. To see that $A(\alpha)$ is closed, observe that if $t_1 \in A(\alpha)$ then so is t for all $t > t_1$. Let $t_2 = \inf\{t | t \in A(\alpha)\}$. Choose a small disc D about $\phi(t_2)$. There is a homotopy F_1 of α to the constant map ζ_0 in $S_{t_2+\varepsilon}$. Choose a homeomorphism h of $\bar{D}$ so that $h | \partial D = \text{id}$ and h moves $\phi(t_2 + \varepsilon)$ to $\phi(t_2 - \varepsilon)$ as shown in Fig. 9. h may be extended by the identity to a homeomorphism of S_0. $h \circ F_1(I^2) \subset S_{t_2}$; hence $t_2 \in A(\alpha)$ and $A(\alpha)$ is closed. From the connectedness of $[t_0, \infty)$, it follows that $t_2 = t_0$ and α is null homotopic in S_{t_0}.

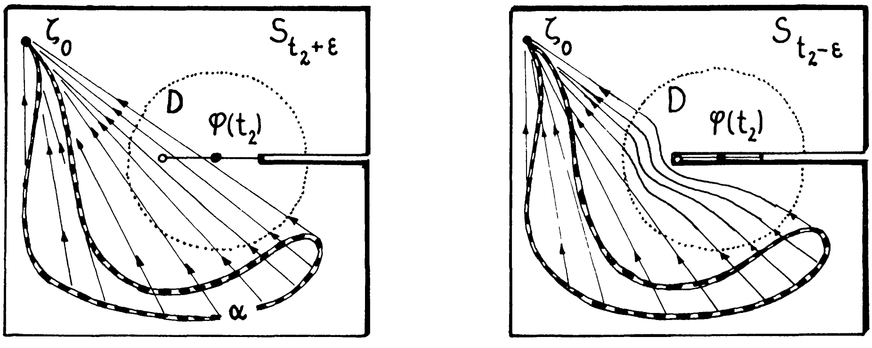


FIG. 9

Proof of Proposition 5.3. The proof of this proposition is modeled after the ideas of Riemann—at least as interpreted by Klein [**13**]. We shall solve a potential problem on a subsurface of S. Then, by studying the level sets and integral curves of the gradient field of that potential, we shall obtain a global topological result. The author begs the indulgence of the knowledgeable reader for presenting several basic results on holomorphic and harmonic functions.

Let f be a function holomorphic in a neighborhood of $0 \in \mathbb{C}$ with $f(0) \neq 0$. Then

$$f(z) = \sum a_n z^n = a_0 + a_k z^k + o(|z|^k)$$

where $o(|z|^k)$ means a function so that

$$\lim_{z \to 0} \frac{o(|z|^k)}{|z|^k} = 0.$$

k is the order of $f(z) - a_0$ at 0.

$\arg(f(z) - a_0) = \arg a_k + k \arg z + o(|z|^k)$. It follows that $\{z \mid \arg(f(z) - a_0) = 0\}$ consists of k analytic arcs which meet at 0 and whose tangents at 0 are equally spaced as a function of $\arg z$. If $k = 1$, then, near 0, $\{z \mid \arg(f(z) - a_0) = 0\}$ is a simple analytic arc. The points where $f'(z) = 0$ are discrete. If $u(z)$ is a harmonic function in a simply connected neighborhood of 0, then it has a harmonic conjugate $v(z)$ there.

$f(z) = \exp(u + iv)$ is holomorphic. 0 is called a *critical point* of u if $f'(0) = 0$. $\{z \mid v(z) = v(0)\}$ consists of k analytic arcs through 0.

The discussion above remains valid for points other than 0 and, indeed, is valid on Riemann surfaces.

Now let S be a simply connected Riemann surface with no divergent curves. Let $\zeta_1 \in S$ and $\overline{N}$ be a closed disc about ζ_1 with analytic boundary. Set $S_1 = S \setminus \overline{N}$. The reader may easily verify that S_1 is homeomorphic to $\dot{S}$, and we must only show that S_1 is simply connected. By Proposition 4.3, S_1 has a Green's function $g(\zeta, \zeta_0)$ for any $\zeta_0 \in S_1$.

As in the proof of Proposition 5.1, we let h_0 be a harmonic conjugate of $g(\zeta, \zeta_0) + \log|z|$ near ζ_0 and let $f_0(\zeta) = z(\zeta)\exp[-(g + \log|z| + ih_0)]$. $f_0(\zeta)$ may be analytically continued along all paths in S_1. Since we do not know *a priori* that S_1 is simply connected, the analytic continuation may not define a function.

Near ζ_0, $f(z(\zeta)) = a_1 z + o(|z|)$ and it follows that f is locally injective near ζ_0 and there is a neighborhood of ζ_0 containing no critical points of g. If g has critical points, let $M = \sup\{g(\zeta, \zeta_0) \mid \zeta \neq \zeta_0 \text{ and } \zeta \text{ a critical point of } g\}$. Otherwise let $M = 0$. Let $D = \{\zeta \in S_1 \mid g(\zeta, \zeta_0) > M\} \cup \{\zeta_0\}$ and $B \subset \mathbb{C}$ be the disc of radius e^{-M} about 0.

LEMMA B.1. *Analytic continuation of f_0 in D defines a holomorphic bijection $f\colon D \to B$.*

Proof. Let $T(\theta)$ be the curve in D starting at ζ_0 along which f_0 continues with constant argument. Since $T(\theta)$ contains no critical points, it is unique and may be parametrized by e^{-g}. Since S has no divergent curves, $T(\theta)$ is a simple arc from ζ_0 to $\partial D = \{\zeta \mid g(\zeta,\zeta_0) = M\}$.

Let $\zeta \in D \setminus \{\zeta_0\}$ and θ_0 be the argument of some analytic continuation f_1 of f_0 to ζ. Then $|f_1(\zeta)| = \exp(-g(\zeta,\zeta_0)$. Let A be the necessarily unique arc through ζ with $\arg f_1 = \theta_0$. Extend A so that it is maximal with respect to being a curve along which f_1 continues with constant argument in D. Since g is monotone on A and D contains no critical points, A is simple and analytic. Thus it is either divergent or contains an arc $T(\theta_1)$ from ζ to ζ_0. The first possibility is ruled out by our initial assumption. Thus $A = T(\theta_1)$ and we may write $\theta(\zeta) = \theta_1$. Thus each point $\zeta \in D \setminus \{\zeta_0\}$ is parametrized in polar coordinates by $(e^{-g}, \theta(\zeta))$, with ζ_0 being the origin. The parametrization is continuous and injective, hence is a homeomorphism of D with B. It follows that D is simply connected and the parametrization is precisely the analytic continuation of f_0.

If $M = 0$, we are done, since $D = S_1$ and f is a homeomorphism of D with Δ. If $M > 0$, we

have

LEMMA B.2. *If $M > 0$, then there is a critical point $\zeta_2 \in \partial D$.*

Proof. First suppose that there is a critical point ζ_2 so that $g(\zeta_2,\zeta_0) = M$. Then arbitrarily close to ζ_2 are points in D and $\zeta_2 \in \partial D$. Otherwise there is a sequence of critical points ζ_n so that $g(\zeta_n,\zeta_0) \to M$. Through each ζ_n there is an arc A_n along which g increases and any continuation of f_0 has constant argument. Since S has no divergent curves, A_n connects ζ_n to ζ_0. For some θ_n, $A_n \cap D = T(\theta_n)$. On a subsequence, $\theta_n \to \theta$ and $T(\theta)$ must have an endpoint ζ_θ. From the local structure of the curves $\arg f =$ constant near ζ_θ, we see that ζ_θ is a limit of critical points. Since critical points of g are discrete in S_1, this is impossible and the Lemma is proved.

LEMMA B.3. *If $M > 0$, then there exists a closed annulus $\bar{A} \subset S$ with piecewise analytic boundary so that $S \setminus A$ is connected. Here $A = \operatorname{Int}\bar{A}$.*

Proof. Using the previous lemma, there is a critical point $\zeta_2 \in \partial D$. Again by the local structure near a critical point, there are at least two curves $T(\theta_1)$ and $T(\theta_2)$ emanating from ζ_2. Choose closed discs $\bar{B}_0$ and $\bar{B}_2$, with centers ζ_0 and ζ_2, respectively, which are defined by $\bar{B}_0 = \{\zeta \mid g(\zeta,\zeta_0) \geqslant \varepsilon^{-1}\}$ and $\bar{B}_2 = \{\zeta \,\|\, g(\zeta,\zeta_0) - g(\zeta_2,\zeta_0)| \leqslant \varepsilon\}$. For ε sufficiently small, the $\bar{B}_i$ are disjoint and contain no critical points. For θ sufficiently close to θ_i, $T(\theta)$ is a curve from B_0 to B_2. Let $\Sigma_i = \cup\{T(\theta) \mid |\theta - \theta_i| < \delta\}$. Σ_i is a strip from B_0 to B_2 for $i = 1,2$. Let $\bar{A} = \Sigma_1 \cup \Sigma_2 \cup B_0 \cup B_2$ as in Fig. 10. For nearly all but finitely many θ_i, $T(\theta_i) \cap \bar{B}_2 = \varnothing$ for ε sufficiently small. Let $\eta \in S \setminus A$ where $A = \operatorname{Int}\bar{A}$. Any analytic continuation f_η of f_0 to a neighborhood of η may be continued along a curve with decreasing modulus and constant argument to ∂S_1. Thus if $g(\eta,\zeta_0) \leqslant M$, the curve will not meet A and η and ∂S_1 lie in the same path component of $S \setminus A$. If $\eta \in D \setminus A$, then η lies in a complementary sector of $\Sigma_1 \cup \Sigma_2$ in D. Choose $T(\theta)$ in that same sector so that continuation along $T(\theta)$ with decreasing modulus and constant argument does not lead us into B_2. We may then further continue to ∂B. ζ, $T(\theta)$ and ∂B thus lie in the same path component of $S \setminus A$, and $S \setminus A$ is path connected.

PROPOSITION B.1. *If S is any Riemann surface and $\bar{A}$ is a closed annulus on S with piecewise analytic boundary and such that $S \setminus \bar{A}$ is path connected, then S is not simply connected.*

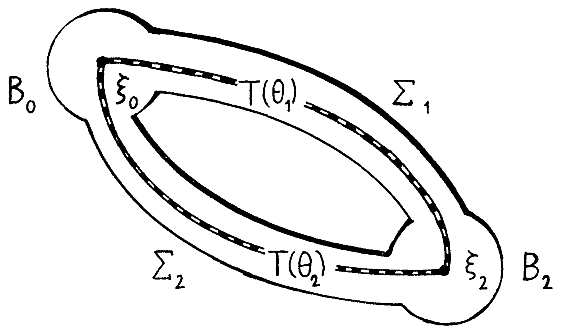


FIG. 10

Proof. $\bar{A}$ has two boundary components F_1 and F_2. Further, there is a simple closed curve α in S so that $\alpha \cap \operatorname{Int}A$ is a simple arc from one boundary component to the other.

Solve the Dirichlet problem in $\bar{A}$ with boundary data $\phi_1|F_1 = 1$ and $\phi_1|F_2 = 0$. The Dirichlet problem in $S \setminus \operatorname{Int}A$ may be solved with boundary data $\phi_2|F_1 = 1$ and $\phi_2|F_2 = 2$. Set $\phi = \exp i\pi\phi_i$ and notice that ϕ is defined on S and continuous. $\phi \circ \alpha$ has winding number $= 1$ about $z = 0$. If α is null homotopic in S then $\phi \circ \alpha$ is null homotopic in $\partial\Delta$. The latter is impossible; hence S is not simply connected.

The research for this paper was partially supported by the National Science Foundation and the Alfred P. Sloan Foundation.

References

1. L. Ahlfors, Complex Analysis, 2nd ed., McGraw-Hill, New York, 1966.
2. ______, Conformal Invariants, McGraw-Hill, New York, 1973.
3. L. Ahlfors and L. Sario, Riemann Surfaces, Princeton University Press, Princeton, N. J., 1960.
4. L. Bers, Riemann Surfaces, Courant Institute Lecture Notes, New York, 1958.
5. ______, Uniformization, moduli and Kleinian groups, Bull. London Math. Soc., 4 (1972) 257–300.
6. ______, On Hilbert's 22nd Problem, Proceedings of Symposia in Pure Mathematics, Amer. Math. Soc., 28 (1976) 559–609.
7. J. Conway, Functions of One Complex Variable, Springer-Verlag, New York, 1973.
8. M. Greenberg, Lectures on Algebraic Topology, Benjamin, New York, 1966.
9. M. Heins, The conformal mapping of simply connected Riemann surfaces, Ann. Math., 50 (1949) 686–690.
10. D. Hilbert, Mathematical Problems, Bull. Amer. Math. Soc., 8 (1902) 437–479.
11. E. Hille, Analytic Function Theory, Blaisdell, New York, 1959 and 1962.
12. K. Kendig, Elementary Algebraic Geometry, Springer-Verlag, New York, 1977.
13. F. Klein, On Riemann's Theory of Algebraic Functions and Their Integrals, Dover, New York, 1963.
14. F. Klein, Collected Works, vol. 3, Springer-Verlag, Berlin, 1923.
15. P. Koebe, Über die Uniformisierung beliebiger analytischen Kurven, Göttinger Nachr. (1907) 191–210.
16. B. Maskit, A theorem on planar covering surfaces with applications to 3-manifolds, Ann. of Math., 81 (1965) 341–355.
17. ______, The conformal group of a plane domain, Amer. J. of Math., 90 (1968) 718–722.
18. ______, Uniformizations of Riemann surfaces, Contributions to Analysis, Academic Press, New York, 1974, pp. 293–312.
19. ______, On the classification of Kleinian groups: I-Koebe groups, Acta Math. 135 (1975) 249–270.
20. H. Poincaré, Sur l'uniformisation des fonctions analytiques, Acta Math. 31 (1907) 1–63.
21. C. Siegel, Topics in Complex Function Theory, vol. 1, Wiley-Interscience, New York, 1969.
22. M. Tsuji, Potential Theory in Modern Function Theory, Maruzen, Tokyo, 1959.
23. H. Weyl, The Concept of a Riemann Surface, 3rd ed., Addison-Wesley, Reading, Mass, 1955.

BLACK WOMEN IN MATHEMATICS IN THE UNITED STATES

PATRICIA C. KENSCHAFT
Department of Mathematics and Computer Science, Montclair State College, Upper Montclair, NJ 07043

Increased attention has been focused on women in mathematics during the past decade, but when I was invited to speak on Black women in mathematics, I could find only two references—a talk by Vivienne Malone Mayes [1] at the Summer Meeting in Kalamazoo in 1975 sponsored by the Association for Women in Mathematics, and the AWM panel I chaired in Atlanta in January, 1978 [2]. Since then I have collected much information, and this article tells about the American Black women holding doctoral degrees in mathematics, all but two of whom I have talked with in the past three years.

The 1970 decennial census revealed more than 1,100 Black women who reported themselves as mathematicians. In that census 244 said that they were college or university teachers, and, of

Adapted from an invited address given at the annual meeting of the Association of Mathematics Teachers of New England in Springfield, Massachusetts, on November 2, 1979.

Patricia Kenschaft received her Ph.D. with a specialty in functional analysis from the University of Pennsylvania in 1973 under the direction of Edward Effros. Since then she has taught at Montclair State College in New Jersey. She is the author or co-author of three textbooks for nontechnical majors published by Worth Publishers, Inc., and is currently preparing a paper on the life of Charlotte Scott, vice president of the AMS in 1906.—*Editors*

these, 178 said that they had finished at least four years of college; 152 had completed five years or more [3]. I have discovered twenty-one American Black women who earned doctorates in pure or applied mathematics before the end of 1980. I would appreciate hearing of others if they exist.

The first year that a Black woman received a Ph.D. degree in pure mathematics from an American university was 1949. In that year there were two: Marjorie Lee Browne at the University of Michigan and Evelyn Boyd Granville at Yale University. The first American white woman, Winifred Edgerton Merrill, to receive such a degree did so in 1886 from Columbia University [4] and the first Black man, Elbert Cox, in 1925 from Cornell [5].

Georgia Caldwell Smith passed the defense for her Ph.D. degree from the University of Pittsburgh in the summer of 1960 but apparently died before it was conferred upon her in early 1961. Thus Gloria Conyers Hewitt became the third American Black woman to be granted a Ph.D. in mathematics when she received hers in 1962 from the University of Washington. But Argelia Velez-Rodriguez, who became a naturalized American citizen in 1972, received her degree from the University of Havana two years earlier, in 1960. (She was apparently the first Black woman to obtain a doctorate in mathematics at that institution, but she knows of four who have done so since.) Thyrsa Frazier Svager was awarded her Ph.D. degree from Ohio State University in 1965. In 1966 three American Black women received the degree: Eleanor Green Jones from Syracuse University, Vivienne Malone Mayes from the University of Texas at Austin, and Shirley Mathis McBay from the University of Georgia.

In 1967 Geraldine Darden received a doctorate in mathematics from Syracuse University, and in 1968 Mary Lovenia DeConge received hers in mathematics and French from St. Louis University. Etta Zuber Falconer completed her degree at Emory University in 1969. In 1971 Dolores Richard Spikes won hers at Louisiana State University. In 1974 Elayne Arrington-Idowu received her doctorate from the University of Cincinnati, Rada Higgins McCreadie from Ohio State University, and Evelyn Patterson Scott from Wayne State University.

There then appears to have been a gap until Fern Hunt was awarded a doctorate from New York University in 1978 and Fannie Gee from the University of Pittsburgh in 1979. In 1980 Frances Sullivan received hers from the City University of New York, Suzanne Craig from the University of Southern California at Los Angeles, and Sylvia Trimble Bozeman from Emory University.*

During my inquiries, I have discovered about a dozen Black women with doctorates in mathematics education, and I feel sure there are more. These women have been leaders not only in teaching and administration, but also in organizing programs and writing. Two of them, Ethel Turner and Joella Gipson, were the author [6] and co-author [5], respectively, of books that contain a list of Blacks who have received doctorates in mathematics and mathematics education. Reference [5] was especially helpful in the preparation of this article because it contains data of the type that appears on a résumé. Both are somewhat out of date, and neither contains anecdotal information or personal views such as those in this article.

Most Black women engaged in mathematical careers do not have doctorates. Their number is probably larger than the number with doctorates would indicate because the economic disincentives for graduate education tend to be greater for Blacks than for whites. I arbitrarily restricted myself to those with doctorates only because the number was sufficiently small for me to interview and they were the ones most interesting to readers of this MONTHLY.

Marjorie Lee Browne was born and raised in Memphis, Tennessee. Her father had attended college for a couple of years and was known as a whiz at mental mathematics [7]. He shared his enthusiasm with his two children and she told me, "I always, always, always loved mathematics!" [9]. With his Civil Service job as a railway postal clerk, he was much better off than most Blacks around him and he felt keenly his obligation to share his good fortune with others, including his

*I have been told that Carolyn Moncure, who received an M.A. in mathematics from Fisk University in 1956, later received a doctorate in mathematics, but I have not been able to locate her.

own children. He enrolled his daughter in LeMoyne High School, a private school started after the Civil War by the Methodist and Congregational Churches to educate Negroes. She reached college age during the Depression, but by a combination of scholarships, jobs while a student, and borrowing, she was able to attend Howard University, from which she graduated in 1935.

For a while she taught at a New Orleans private secondary school for Blacks, Gilbert Academy. Then a neighbor in Memphis told her that the fees at the University of Michigan were in the range she could afford. She enrolled at that university, receiving her M.S. degree in 1939. She then joined the faculty of Wiley College in Marshall, Texas, and began working toward her doctorate during the summers. Eventually she took a leave from Wiley College to return to the University of Michigan, and in 1949 she received her Ph.D. degree. Her thesis, "On the One Parameter Subgroups in Certain Topological and Matrix Groups," was written under the supervision of G. Y. Rainich.

She taught at North Carolina Central University from 1949 to 1979 and for 25 years was the only person in her department with a Ph.D. in mathematics. She taught 15 hours a week, both undergraduate and graduate courses. In addition, she supervised the writing of ten master's degree theses and was department head from 1951 to 1970. She was the principal author of a proposal for a grant from IBM which gave NCCU its first electronic digital computer for academic computing, and she supervised its installation.

Her summers were occupied with teaching secondary teachers. Under her leadership NCCU became the first predominantly Black institution in the United States to be granted funds for an NSF Institute for secondary teachers of mathematics. She directed the mathematics section of these Institutes for 13 years and wrote four sets of Lecture Notes for their use: "Sets, Logic, and Mathematical Thought," 1957; "Introduction to Linear Algebra," 1959; "Elementary Matrix Algebra," 1969; and "Algebraic Structures," 1964.

In August 1955, her paper "A Note on the Classical Groups" was published in this MONTHLY. She had numerous special appointments with the NSF, such as serving on the Advisory Panel on Undergraduate Scientific Equipment Program in 1966, 1967, and 1973. In the year 1968–69 she was a Faculty Consultant in Mathematics for the Ford Foundation. At various times she won grants for postdoctoral study at Stanford University, Columbia University, UCLA, and Cambridge University.

Dr. Browne died on October 19, 1979, at the age of 65. Less than two weeks before, she told me in a telephone conversation: "If I had my life to live again, I wouldn't do anything else. I love mathematics. I do have plenty of friends, and I talk with them for hours at a time. But I also like to be alone, and mathematics is something I can do completely alone."

The scholarship committee soliciting funds for the Marjorie Lee Browne Trust Fund at NCCU has written, "Her thoroughness, wisdom, vision, productive powers, and demands for excellence and rigor in the classroom profoundly influenced not only the academic growth and development of countless students, but also their aspirations to achieve and succeed in many professions after leaving the University ... It was not uncommon in recent years for her to make substantial financial contributions to able students who needed assistance in acquiring an education."

Evelyn Boyd Granville grew up in Washington, D.C., and attended the academic segregated high school there, Dunbar High School. This school had many fine teachers who urged their students to apply to the best colleges. Among them were mathematics teachers Ulysses Basset, a graduate of Yale, and Mary Cromwell, a graduate of the University of Pennsylvania.

Granville won a partial scholarship to Smith College; her mother and an aunt made many sacrifices in order to meet the rest of her bills. In 1945 she graduated from Smith College summa cum laude and with election to Phi Beta Kappa. From Smith she went to Yale, where she worked with Einar Hille, writing a Ph.D. dissertation "On Laguerre Series in the Complex Domain." Although she and Marjorie Lee Browne are linked in history by receiving their doctoral degrees in 1949, they never met.

In 1950, after a year of postdoctoral research at NYU, she applied for an academic position

and had several interviews. A faculty member at one institution where she was interviewed told me recently that, when the hiring committee discovered she was Black, they merely laughed at her application and never considered her for the job, even though she had a Ph.D. degree from Yale under Hille. I report this institution anonymously because I suspect that its behavior was by no means unique. When I mentioned the incident to Dr. Granville in 1980, she told me that she remembered the interview, but had not known before that race had been a factor in her not receiving a job offer.

She taught at Fisk University from 1950 to 1952 and had a lifelong impact on several of her students. At least two of them, Vivienne Malone Mayes and Etta Zuber Falconer, later received doctorates in mathematics themselves. Mayes has written, "I believe it was her presence and influence which account for my pursuit of advanced degrees in mathematics." Granville spent 16 years in government and industry. The first three years were with Diamond Ordnance Fuze Laboratories, analyzing mathematical problems arising in the development of missile fuzes. From 1956 to 1960 she worked on the formulation of orbit computations and computer procedures for Project Vanguard and Project Mercury at IBM. The following two years were spent at Space Technology Laboratories doing research on space trajectories. She then spent a year with North American Aviation Space and Information Systems Division as a research specialist on the Apollo project in the areas of celestial mechanics, trajectory and orbit computation, numerical analysis, and digital computer techniques. From 1963 to 1967 she continued her work in these same areas, again at IBM.

Since 1967 she has taught at California State College in Los Angeles, where she is now a full professor. She is coauthor of the book *Theory and Application of Mathematics for Teachers*, published in 1975 by Wadsworth Publishing Company. A second edition appeared in 1978. She has had numerous special appointments, including serving on the Psychology Examining Committee of the Board of Medical Examiners for the State of California from 1963 to 1970 and as a member of the Center for the Improvement of Mathematics Education since 1975.

She has written to me that, looking back, she feels that she has a "very rich life" and attributes her "blessings" to a supportive family, a Black community and high school in Washington, D.C., that "recognized and applauded excellence in education," generous scholarship and fellowship aid, and the fact that during her time Blacks could enter the job market. Her own perspective in using these opportunities is suggested by her statement, "I always smile when I hear that women cannot excel in mathematics."

Lee Lorch joined the faculty of Fisk University at the same time as Granville and was head of the mathematics department there from 1950 to 1955. After the United States Supreme Court's decision of May 17, 1954, declaring segregation in public education to be unconstitutional, Lorch and his wife attempted to enroll their daughter in the school nearest their home, a Black school across the street. The Nashville School Board turned them down, the four white members against the one Black. Shortly thereafter he was subpoenaed by the House Committee on Un-American Activities and subsequently was dropped from the Fisk faculty, by a split vote of the predominantly white board of Trustees, without charges or trial. During the years 1955–58 he taught at Philander Smith College in Little Rock, Arkansas, another Black university. (One of his students there, Frank James, became the only graduate of that institution to obtain a Ph.D. in mathematics.) When the school integration crisis erupted in Little Rock in 1957–58, the Lorches were active supporters of the Black students. Thereupon Senator James Eastland of Mississippi summoned Lee's wife, Grace Lonergan Lorch, before his United States Senate Committee. This and related events essentially brought the Lorches' time in the United States to an involuntary end. Three of the women mentioned in this article (Falconer, Hewitt, and Mayes) studied under Lorch during his five years at Fisk; Mayes summarized his impressive accomplishments there in an earlier article in this MONTHLY [**8**]. One might sadly wonder if proportionately more Black women (15 of them) might have received doctorates in mathematics by now if Lorch had been permitted to spend the past 25 years at Fisk.

Gloria Conyers Hewitt's first two years at Fisk coincided with Lorch's last two. During her senior year, she discovered that he had recommended her to two graduate schools and suggested that they encourage her to pursue a doctoral program with them. She reports, "As he had taught me for such a short time, it amazes me to this day that he felt I could handle graduate school in mathematics... the thought of entering graduate school in mathematics never crossed my mind; I never knew it crossed his until I heard of his recommendations" [**8**]. Five years later she received her doctorate from the University of Washington. When she entered graduate school, she felt that her mathematical education had ended with the departure of Lorch at the end of her sophomore year; she had so much to make up. She gives credit for her ability to do this to her family's encouragement and to that of her fellow students. She said, "It was almost as if they had gotten together and decided that I should get the degree" [**9**]. At first she was afraid to ask the other graduate students questions. They would come to her office and say, "Isn't this an interesting problem? Let me show you how interesting it is!" She remembers that Carl Stromberg was often in her office when she first arrived in the morning, ready to show her the problem of the day. She also feels especially grateful to Kenneth Ross and Robert Phelps. One day the latter said to her, "The trouble with you is that you don't know trigonometry. I am going to teach it to you." She listened for two hours and then she understood trigonometry. There were eight or ten students who helped her significantly, all of them white males. Eventually, of course, she felt free to approach them with questions of her own. When they went to a tavern in the evening, she went along. If they went to the best restaurant in Seattle, she went too. She also appreciates the help of her thesis advisor, Richard Pierce, "whose faith in me remained even when my own faltered" [**9**]. In 1962 she received her Ph.D. degree with the thesis, "Direct and Inverse Limits of Abstract Algebras."

Hewitt's first position was at the University of Montana, where she has been a full professor since 1973. She held an NSF Postdoctoral Science Faculty Fellowship at the University of Oregon in 1965–66 and was on sabbatical leave at Case Western Reserve University during the year 1980–81. Her papers include "The Existence of Free Unions in Classes of Abstract Algebras," *Proc. Amer. Math. Soc.*, 14 (1963) 417–422, and "Limits in Certain Classes of Abstract Algebras," *Pacific J. Math.*, 22, no. 1 (1967). She has been active in the AAAS, has been a Visiting Lecturer for the MAA, and has read many proposals for the NSF and HEW.

Vivienne Malone Mayes was a junior at Fisk when Lorch arrived. She became his grader, and at his urging remained after her graduation in 1952 for graduate work. She received her M.A. from Fisk in 1954. Stimulated by Lorch to do research, she later enrolled in a doctoral program at the University of Texas at Austin, after teaching for seven years at Paul Quinn College.

Mayes grew up in Waco, Texas, where she described the public schools as "strictly separate and strictly unequal." Although she was "victimized educationally" by the segregation, she has also written, "In every Black school I've attended there's always been at least one Black woman teacher or professor with whom I could identify... No difference was made between boys and girls... Every girl expected to work. Her hope was that through education she could escape the extremely low-paying jobs designated for Black women... Girls held the majority in my upper-level math classes at Fisk... After Fisk, I taught at two Black colleges. In both instances, girls outnumbered boys in every class" [**1**]. Her experiences at graduate school were in stark contrast. In her first class, "I was the only Black and the only woman. For nine weeks thirty or forty white men ignored me completely... It seemed to me that conversations before class on mathematics between classmates quickly terminated if it appeared that I was listening... My mathematical isolation was complete."

Mayes could not become a teaching assistant at the University of Texas in Austin because she was Black. There was even one professor who would not let Blacks attend his classes. She could not join her advisor, Don E. Edmondson, and other classmates to discuss mathematics over coffee at Hilsberg's Cafe because she was Black. Sometimes she caught snatches of their conversations as they crossed the picket lines. Only after the law changed to require Hilsberg's Cafe to serve Blacks

did she notice that women were rarely included in these conversations no matter what their color. After several courses friendships did begin to develop, but throughout her graduate years there was only one other woman in most of her classes. They became good friends.

Despite these obstacles she received her doctorate in 1966 with the dissertation, "Some Steady State Properties of $(\int_0^x f(t)\,dt) \div f(x)$," published in the *Proc. Amer. Math. Soc.*, 22, no. 3 (1969) 672–677. Her résumé lists six publications since then of which the most recent is "Some properties of the Leininger generalized Hausdorff Matrix," in the *Houston J. Math.*, 6, no. 2 (1980) 287–299.

Since receiving her doctorate, Mayes has been on the faculty at Baylor University in Waco, where she is now a full professor. In 1971 she was elected "Outstanding Faculty Member of the Year" by the Student Congress. She has given many speeches, including several on "Audio Tutorial Approaches to Pre-Calculus." Active in professional organizations, she served three terms on the MAA national committee on the High School Lecture Program and was on the Program Committee of the MAA for its national meeting in Dallas in January 1972. She was elected Director-at-Large for the Texas Section of the MAA for 1973–74.

Georgia Caldwell Smith received her bachelor's degree from the University of Kansas about 1929, was a member of the faculty at Spelman College, and completed the work for a Ph.D. degree in mathematics at the University of Pittsburgh in the summer of 1960. She wrote "Some Results on the Anticenter of a Group" under the direction of Norman Levine. Shortly thereafter she died, apparently before actually being awarded the degree in January 1961.

Argelia Velez-Rodriguez grew up in Havana, her birthplace, where she attended Catholic elementary and secondary schools which provided consistently good teachers. She won a medal in a competition about fractions in third grade and ever after was the acknowledged leader in her class in mathematics. Her interest in mathematics has continued ever since. "There is no racial problem in Cuba," she reports, but adds that discriminatory practices did exist when places were owned or controlled by citizens of the United States [**9**]. As a child she was aware of these situations but was not affected by racism herself. Socioeconomic discrimination existed in pre-Castro Cuba, but since her father had a steady government job his five children were able to obtain an education. Family expectations were high. (This has continued into the next generation; her son is now in medical school and her daughter recently received a bachelor's degree in Industrial Engineering from Stanford University.)

When Velez-Rodriguez was an undergraduate at the University of Havana, most of her mathematics and science instructors were women, all of whom had doctorates. There were no Black women in mathematics but there was one in science. These women encouraged her to continue her own education. The main problem deterring her from graduate study was the fact that there were more qualified mathematicians in Cuba at that time than there were job openings. All the other students in her graduate school classes were white, but she felt no discrimination. In 1960 she received her Sc.D. degree, writing her dissertation, "Determination of Orbits Using Talcott's Method," under Manuel Rabina. Her fields are differential equations and astronomy. Since 1960 she has had five expository mathematical publications, including *Functions and Analytic Geometry*, the Curriculum Resources Group, Institute for Services to Education, Massachusetts, 1969, and *Las Mathematicas: Lenguaje Universal* (English-Spanish version), Bilingual Materials Development Center, Fort Worth, Texas, 1977.

Meanwhile, her son was born in 1955 and her daughter in 1959, the year of the Cuban revolution. Her son entered a Catholic school, and she tried to live under the new government while she completed her own education. But when her son's school was taken over by the government and she realized he would be educated as a Communist, not a Catholic, the situation seemed intolerable. With the help of a Catholic agency, she obtained a visa so that he could be educated in the United States. She and her baby daughter were allowed to accompany him, and the three left Cuba in 1962; but her husband was forced to remain there for three more years. She had been free of racism and sexism in Cuba, but was soon made keenly aware of the "tension and pressure" they cause in this country. She acknowledges that in her case the usual problems faced

by Blacks and women were compounded by a language change. Nevertheless, she has taught in several American schools and was chairperson of the Department of Mathematical Science of Bishop College in Dallas, Texas, from 1975 to 1978.

She has held more than ten other administrative positions, including Associate Director of the Cooperative Doctoral Program in Mathematics Education between the University of Houston and Bishop College, in 1973–75, supported by HEW, and the Director and Coordinator of the Project on Computer-Related Mathematics for Senior High School Teachers, in 1977–78, sponsored by the NSF. Her résumé lists almost 30 committees and task forces on which she has served and 10 special panels and speeches. In 1976 she received a travel grant from the National Research Council to attend the Third International Congress on Mathematics Education in Karlsruhe, Germany. In 1979 she took a leave of absence from Bishop College so that she could become a program manager with the Minority Institutions Science Improvement Program (MISIP) in Washington. It was then under the auspices of the NSF, but was transferred to the new Department of Education in the spring of 1980, about which time she was promoted to program director.

Thyrsa Frazier Svager is a graduate of Antioch College. She was employed for one year each as a statistical analyst at Wright-Patterson Air Force Base in Dayton, Ohio, and as an instructor at Texas Southern University in Houston, and since 1954 has been a member of the faculty of Central State University in Wilberforce, Ohio. Both her M.A. and Ph.D. were awarded by Ohio State University in Columbus, the latter in 1965 under the direction of Paul Reichelderfer. Her dissertation was "On Absolutely Continuous Transformations of Measure Spaces." Her other publications include "On Strong Differentiability," whose abstract appeared in this MONTHLY in 1967, and two books: *Modern Elementary Algebra Workbook*, Wm. C. Brown, 1969, and *Essential Mathematics for College Freshmen*, Kendall-Hunt, 1976. Her current research interest is computer applications in number theory. Since 1966 she has been a professor and department chairman at Central State.

Shirley Mathis McBay was born in Bainbridge, Georgia. Her first two degrees were in chemistry, a B.A. from Paine College in Augusta, Georgia, in 1954 and an M.S. from Atlanta University in 1957. She taught at Spelman College and was head of the Division of Natural Sciences there. In 1966 she received a doctorate in mathematics from the University of Georgia in Athens. Her supervisor was Thomas R. Brahana; her dissertation was entitled "The Homology Theory of Metabelian Lie Algebras." After administering NSF programs for several years, in 1980 she became the Dean of Students at the Massachusetts Institute of Technology.

Eleanor Green Dawley Jones was the second of six children of a letter carrier in Norfolk, Virginia. She was educated in completely segregated schools and the only whites she knew as a child were the priests and nuns of her parish. When she graduated from high school at the age of 15, she won a scholarship to Howard University. There she studied under Elbert Cox, the first Black American to receive a Ph.D. in mathematics, as well as several other Black men with doctorates. She received her B.S. in 1949 and her M.S. in 1950, both in mathematics. For a while after her graduation she taught in high school; in 1955 she became an instructor at Hampton Institute. The Black men she had known with doctorates served as role models, and it occurred to her that she should work for a Ph.D. degree. In 1962 she left Hampton Institute to work for a doctoral degree in mathematics at Syracuse University. At that time no Black people were allowed to pursue doctoral studies in any academic discipline in Virginia, but the state would pay tuition and travel costs of Black citizens who went out of the state for graduate study. Jones by then had two small sons, whom she took with her to Syracuse. There she earned her own living and that of her children while obtaining her doctorate, which she did in 1966. Her thesis advisor was James Reid and her thesis topic was "Abelian Groups and Their Endomorphism Rings and the Quasi-Endomorphism of Torsion Free Abelian Groups"; part of this was published as "A Note on Abelian p-Groups and Their Endomorphism Rings" in this MONTHLY, May 1967. Her paper "$4^x + 4^y + 4^z =$ a Square" appeared in March 1969, also in this MONTHLY. After obtaining her doctorate, she returned to Hampton Institute for a year as an associate professor and then obtained a position at Norfolk State College, where she is now a full professor.

While Jones was teaching at Hampton Institute, before she earned her doctorate, she was assisted by an upper-class grader by the name of Geraldine Darden. She, too, had grown up in Virginia and had attended all-Black elementary and secondary schools. She was a good student, but the only career that occurred to her was that of a high school teacher; no one had pointed out other career possibilities. So after she graduated from Hampton Institute in 1957, she began teaching in a high school. Six months later she thought she had made a mistake. Fortunately for her, that was a time when there was a renewed national interest in mathematics. She applied for and received a grant to attend one of the summer institutes Marjorie Lee Browne was directing at North Carolina Central University.

When she arrived at Dr. Browne's office, she said timidly, "Good morning. I'm Geraldine Darden." Dr. Browne looked up from her desk and said, "Why aren't you in graduate school?" Darden looked around to see if she had interrupted something, but soon the question was repeated, and there was clearly no one else in sight. She paused. "Well, I had to go to work when I finished school." "So you could earn that big car you are driving out there?" "No, I come from a big family and I needed a car, but I had to go to work because I had little brothers and sisters at home and they needed help." From then on Darden was continually encouraged by Browne. As a result she went to the University of Illinois the following year and received an M.S. in mathematics education in 1960 [**2**].

With this she obtained a job at Hampton Institute and discovered that she liked teaching at the college level. She decided to continue her graduate work, but the president of Hampton Institute would give her a leave of absence only if she enrolled in a Ph.D. program. Since Illinois would give only two degrees to any one person, and she wanted the security of obtaining a Master's degree in mathematics before working toward a doctorate, she went to Syracuse University. She doubted that she was capable of Ph.D. level achievements, but she worked hard.

At the end of the first year she went to the office of Arthur Sagle to pick up her final examination in abstract algebra. He said, "You did well. You came in third." She was astounded because she knew that many of the other students were very good. "Are you sure you know who I am?" He laughed, for she was the only woman and the only Black in the class. He then asked her what degree she intended to pursue, and when she expressed doubts about her ability to get a Ph.D. he responded, "You can do it!" She remembers that as the time she became convinced she could get a doctorate [**2**]. She wrote her dissertation, "On the Direct Sums of Cyclic Groups," under the supervision of James Reid, and received her Ph.D. degree in 1967. She is now department head at Hampton Institute.

Mary Lovenia DeConge was born and raised in Louisiana, where she was inspired by a high school mathematics teacher. At the age of sixteen she entered the congregation of the Sisters of the Holy Family, taking her vows some years later. From 1952 to 1955 she taught in the parochial elementary schools in the Baton Rouge and Lafayette Dioceses. She then entered Seton Hill College in Greensburg, Pennsylvania, from which she received a B.A. in mathematics and French in 1959. She taught at the Holy Ghost High School in Opelousas, Louisiana, from 1959 to 1964, except for 1961–62, when she held an NSF grant that enabled her to earn an M.A. in mathematics from Louisiana State University.

In 1964 she began full-time graduate study at St. Louis University, and four years later she was awarded a Ph.D. degree in mathematics with a minor in French. This involved taking specified courses in French and then passing the oral comprehensives in that discipline. Her dissertation, "2-Normed Lattices and 2-Metric Spaces," was written under the direction of Raymond Freese and her degree was awarded in 1968.

After teaching at Loyola University from 1968 to 1971, she joined the faculty of Southern University in Baton Rouge, Louisiana, where she still is. There she has directed theses for the degree of Master of Teaching in Mathematics. Her paper "D_2 Lattices" appeared in the *AMS Notices* of January 1971.

Etta Zuber Falconer was another student of both Granville and Lorch at Fisk University. She obtained her A.B. in 1953 and, encouraged by them, earned an M.S. in mathematics from the University of Wisconsin a year later. She taught at both the high school and college levels and in

1965 joined the faculty of Spelman College [5]. Her doctoral work at Emory University was supervised by Trevor Evans. In 1969 she completed her Ph.D. degree, writing "Quasigroup Identities Invariant Under Isotopy." Two papers resulted from this dissertation: "Isotopy Invariants in Quasigroups" in the *Transactions of the AMS*, 151 (1970) 511–526, and "Isotopes of Some Special Quasigroup Varieties" in *Acta Math. Acad. Sci. Hungar.*, 22 (1971), 73–79. After McBay left Spelman, Falconer became head of the Division of Natural Sciences there and currently holds that position.

Dolores Spikes was born and grew up in Baton Rouge. She graduated from Southern University in that city and then received her M.S. degree from the University of Illinois in 1958. She did not consider pursuing a Ph.D. degree at that time, mainly because she had no personal knowledge of Black Ph.D. mathematicians. She wrote to me, "I honestly believe that the question of role models weighed heavily here. Of course, this was prior to graduate education's being (comparatively) more open to Blacks in southern institutions of higher learning" [9]. Instead, she taught at Southern University for two years and then at Mossville High School in Westlake, Louisiana, for four years. When her husband accepted a position in Baton Rouge in 1965, she began to ponder the possibility of returning to school. She had a young daughter, and Louisiana State University, also in Baton Rouge, seemed to offer the least disruption to her family life. She was able to study full time while working on her doctorate; first she held an NSF Science Faculty Fellowship and then a Ford Foundation Fellowship.

In 1971 she completed her dissertation, "Semi-valuations and Groups of Divisibility" under the direction of Jack Ohm. She believes she thereby became the first graduate of Southern to have received a Ph.D. degree in mathematics, and she was indeed the first to do so at L.S.U. In the 1978 Atlanta panel [2] she said, "I am not proud of that fact. I regard it as a shame—a blight on the state of Louisiana and on education in general." She is now a professor of mathematics at Southern University.

Evelyn Patterson Scott grew up in Birmingham, Alabama, and was greatly influenced by her high school geometry teacher, the late William S. Peterman. After she graduated from high school, he helped her obtain a partial scholarship to Alabama A & M University in Normal. After earning a bachelor's degree there, she taught high school in Birmingham and other places, meanwhile earning a master's degree in mathematics from Atlanta University. Later she attended Wayne State University, from which she received her doctorate in 1974. Her advisor was Chia K. Tsao and her thesis title, "An Alternative Bayesian Model." Since then she has been an operations research analyst, first with the Center of Naval Analysis and more recently with the General Services Administration.

Elayne Arrington-Idowu spoke in the 1978 Atlanta panel of the disadvantages of growing up Black and female in the North. In her suburban Pittsburgh town she was not permitted to be a cheerleader or a drum majorette—and certainly not an angel in the Christmas play. It is with obvious joy that she tells about her two daughters who won national contests in baton twirling. When she was valedictorian of her high school class, she was not allowed to give the valedictory address; that was the only year the class president gave the graduation speech.

Her mother worked hard in a restaurant and there was little money for college. Thus they were overjoyed when she received a letter saying that on the basis of her high school grades and college board scores she had ranked first for a company scholarship, which would replace the tuition scholarship she had been awarded by the University of Pittsburgh and in addition pay her living expenses, including funds for books. A few days later she was informed that the company refused to give its scholarship to a female; thus she would receive only the University tuition scholarship. She has written, "First of all I felt resentful—I resented the white male who received 'my scholarship,' as I thought of it. I had classes at the university with him and often heard him boast about the great scholarship that he had 'won.' Secondly, I felt compelled to prove that I could do anything that the male students could do (often collectively). They were not friendly and I didn't ask them any questions, lest they think that I was not capable of doing my own work. In effect, I was isolated, and it was me against all of them... the worst effect of my undergraduate experiences was the lack of intellectual exchange with my peer group" [2].

When she graduated from the University of Pittsburgh, her class standing was not listed because women were not included; she had to compare her average with the men on the list to discover where she stood. After college she was an aerospace engineer at Wright-Patterson Air Force Base for seven years, and then she enrolled in graduate school at the University of Cincinnati. There she was told by her prospective advisor that she should spend an extra year preparing for the preliminary exams because although the general failure rate was 50 percent, the rate among "housewives like you" was 98 percent. She was startled to think of herself as a "housewife," but was persuaded to spend the extra year. In 1974 she received her doctorate with a thesis entitled "The p-Frattini Subgroup of a Finite Group" under the direction of Donald B. Parker. Later she extended this research; her paper "p-Saturated Formations" was published in the *Israel Journal of Mathematics* in November 1978. She is now a member of the faculty of the University of Pittsburgh.

Rada Higgins McCreadie grew up in Columbus, Ohio, and received a good education in the public schools there. During tenth grade she was admitted to an NSF program in which she studied mathematics at Ohio State University during the three summers following her tenth, eleventh, and twelfth grades. At that time her mother, who has a doctorate in education, was teaching at Ohio State. In the following two years McCreadie was an undergraduate at Sacramento State College while her mother taught there. Then they both went to Miami University in Coral Gables, Florida, where McCreadie received her bachelor's degree in 1969.

In 1974 she received her Ph.D. degree from Ohio State University with a dissertation "On the Asymptotic Behavior of Certain Sequences." Her advisor was Ranko Bojanic, who told me she was "one of the best students I have had." From 1974 to 1976 she was a visiting assistant professor at Ohio State and then she spent a year at Howard University. Since then she has been in Holland with her husband. (She and Smith are the two on this list with whom I have not spoken.) During the past few years she has been a research assistant at Delft University in Delft, Holland.

Fern Hunt grew up in New York City and attended the public schools there. At age ten she decided that she wanted to be a scientist; her mother was extremely supportive of her academic interests. She was further encouraged by her science teacher, Charles Wilson, at LaSalle Junior High School. He was a Black man with a master's degree in chemistry from Columbia University and the first person she encountered who used scientific and mathematical reasoning at a high level. She believes that half of his former students later earned college degrees. He suggested that she apply for a Science Honors Program at Columbia that ran summers and Saturdays through her high school years. Her primary interest changed from science to mathematics as a result of a course in this program.

She attended the Bronx High School of Science and then Bryn Mawr College. After her graduation from college she worked for a short time doing mathematical modeling at Abt Associates in Cambridge, Massachusetts, and then became a lecturer at City College in New York City while pursuing graduate work at the Courant Institute. During this time she met Frances Sullivan, who encouraged her to continue her studies and who found the inspiration in Hunt's accomplishments to begin her own graduate program. Hunt met "many good people" at the Institute and she feels fortunate to have had Frank Hoppensteadt as a thesis advisor. In 1978 she completed her degree in mathematical biology with a dissertation, "Genetic and Spatial Variation in Some Selection Migration Models." She then joined the mathematics department at Howard University, where she continues to apply differential equations to population genetics.

Fannie Gee grew up in Carthage, Mississippi, where there were only 42 students in her segregated high school class. They all had to take the same courses, and only three years of mathematics were offered. Worse, the last year, Algebra 2 was taught by the shop teacher who was also the assistant principal, and he often had to miss class without warning. However, the students had great respect for their teachers, and she feels especially grateful to her ninth grade algebra teacher, who is now a member of the State Legislature. During her high school years mathematics seemed "like a game."

When she entered Alcorn State College, also in Mississippi, she found she was lacking in

mathematical preparation, but was motivated to catch up. There the faculty "pushed you," and after she was graduated in 1967 she was well prepared to begin work on an M.A. in mathematics at DePauw University in Indiana. The social adjustment was eased by a classmate from Alcorn who also enrolled in DePauw to study mathematics. After receiving her M.A. in 1969, she taught at T. J. Rogers Junior High School in Houston, Texas, and then returned to Alcorn as an Instructor. The administration at Alcorn urged her to pursue a doctorate, and so she enrolled in the University of Pittsburgh in 1973.

Her first three years at Pittsburgh were largely financed by federal funds funneled through a consortium of predominantly Black institutions. Fortunately, there were others in the same program at the university, because she found that "you might not hear a 'hello' for days except from the ones you knew." The University helped support her last two years of doctoral study and Alcorn contributed "spending cash." In 1979 she completed her dissertation, "A Characterization of LaGrangian Groups," supervised by W. E. Deskins, and returned to the faculty of Alcorn State College.

Frances Sullivan grew up in Orangeburg, South Carolina, and was graduated from the state college there in 1965. During her college years she helped organize the local nonviolent protests of the civil rights movement. Her main interest was in improving the quality of education offered to Black people in South Carolina. Later she attended the University of South Carolina, from which she received an M.S. degree in mathematics in 1968.

She then taught at several different places, including the SEEK program (Search for Education and Elevation Through Knowledge) at City College in New York. She later joined the mathematics department at City College and began a doctoral program at CUNY. During her last four years as a student she won a variety of fellowships that enabled her to pursue her education full time. In 1980 she completed her Ph.D. thesis, "Wreath Products of Lie Algebras," under the direction of Gilbert Baumslag and then joined the faculty of Jackson State University in Mississippi.

Suzanne Craig was born and grew up in Chicago and attended the University of Chicago High School. She remembers especially the encouragement of her white fourth grade teacher, Miss Shaughnassy, and a high school mathematics teacher, Miss Johns. She graduated from the University of Michigan in 1969 having been elected to Phi Beta Kappa. For four years she taught high school mathematics in Detroit and then returned to the University of Michigan, from which she received a master's degree in mathematics in 1975.

Her doctoral work in topological dynamics and differential equations was done under the direction of Robert Sacker at the University of Southern California at Los Angeles; her dissertation title was "Strong Trichotomies and the Splitting Index for Linear Differential Systems." In June 1980, she received her Ph.D. degree, and since then has been a senior engineer at the Jet Propulsion Laboratory in Pasadena, California.

Sylvia Trimble Bozeman is a native of Camp Hill, Alabama, and went through the segregated public school system there. She reports that her parents and teachers had similar educational values and instilled in her a positive attitude about herself, a feeling that "I could do anything I wanted to do." She was especially inspired by her high school geometry teacher, Frank Holley. She received her undergraduate degree from Alabama A & M in Huntsville in 1968 and went directly to Vanderbilt University, from which she received her master's degree in January 1970. This graduate study was financed by the Southern Fellowship Fund. Her two children were born subsequently and she taught part time, first at Vanderbilt and then in the Upward Bound Program at Tennessee State University. In 1974 she joined the faculty of Spelman College, where she is currently employed. When she began her graduate work at Vanderbilt, joining for the first time in classes with white students, she discovered that she was missing some undergraduate courses, but found she could compete well in courses such as topology where the other students were just beginning too. Most notably, she had never had a course in linear algebra, and she was encouraged to take it in an undergraduate class at Vanderbilt. She refused, saying she would teach it to herself, and she did.

She was a full-time student at Emory University from 1976 to 1979, working toward a doctorate in mathematics. During her first year she was exposed to functional analysis in a "great" course given by John Neuberger that kindled her interest in the subject. She notes that he seemed to have a hobby of encouraging women in mathematics, and, in particular, he gave her a lot of support in being both a mother and a mathematician. Later she took a course in numerical analysis under Luis Kramarz. She told me, "He is so enthusiastic about the mathematics he does that I decided to work with him." She says that her most ardent supporter throughout her career, however, has been her husband, who also holds a Ph.D. degree in mathematics. She received her doctorate in August 1980 with a dissertation on "Representations of Generalized Inverses of Fredholm Operators."

What do these women have in common besides mathematical talent? Certainly they all have emotional stamina. In each case there seems to be someone in the family who believed that she was very special and that it was worth sacrificing for her education. All the women apparently remember at least one supportive teacher along the way—someone who knew mathematics and told the young woman that she could be a mathematician.

One might ask why only twenty-one Black women in our country have thus far been able to earn a doctoral degree in mathematics. Despite a variety of efforts to eliminate it, discrimination continues, often perpetrated unconsciously by people of good will. Some of these women feel that the educational hurdles for Blacks are worse now than a decade ago, especially in the large Northern cities. This means that although large numbers of children are responding with interest to the increased publicity given to the sciences, the lack of quality education at the primary and secondary levels is preventing them from fulfilling their scientific ambitions.

The results of past discrimination remain; these include, but are by no means limited to, poorer schools in predominantly Black neighborhoods, the need of young educated Black people to support younger students in their own families, and extra administrative duties devolving on those from underrepresented groups in the mathematical community. Often Blacks and women feel a responsibility to help others in their own groups, which takes time and energy that would otherwise be used in vigorously pursuing their own professional careers.

What can be done? Besides encouraging readers of this MONTHLY to support gifted students regardless of race or sex, I feel I can best begin to answer this question by quoting Dolores Spike's statement at the Atlanta [2] panel.*

"I think there is an excellent opportunity for training in the newly emphasized areas of mathematics, such as applied mathematics, in predominantly Black institutions. There the percentage of Ph.D.'s is very low, so this provides a situation where there are able people who could learn in areas where we need more mathematicians.

"What is needed? We need money. But we need more too. Many of these Black women would love to have more training, but they aren't going to leave their families and go to school. I think we have to be innovative in meeting their problems.

"Maybe we can go to them. Maybe we could send visiting specialists, have on-site programs, prepare materials for self-study, and run seminars for these students. I think that this is the way the professional organizations can help these institutions.

"I believe that the mathematical societies should also encourage the direction of more funds from agencies such as the NSF to the kinds of projects that I have enumerated. Another thing I think that these organizations can do is to provide more information to scholars and to college administrators; part of our problem is that these people do not know the plight of Black women mathematicians. We can tell them, but I think that the news coming from a professional organization would have more impact; I think they would be more likely to receive it and perhaps act on it.

*This is slightly edited and checked by her.

"I think that the mathematical organizations could consider a voluntary accreditation program, because I think this would be particularly beneficial to Black institutions. Such institutions will always be on the short end of the funding in states that have predominantly Black schools, but I have noticed that in those subject areas that have accrediting agencies, even with volunteers, they come out a bit better in the distribution of money.

"Lastly, I would suggest that the mathematical community initiate a massive public media campaign. I think we ought to help the public understand what our problems are and what the possible solutions are."

Mathematical talent is both scarce and precious; our society cannot afford to waste it. Possessors of such talent, whatever their race or sex, need encouragement and help if they are to develop fully their abilities and to enter fields previously denied to them. These Black women are role models for those who follow. They should inspire all of us to look for talent wherever it is found and try to provide the needed support.

Many people helped in the preparation of this article by responding to brief telephone calls and letters. Several helped in more time-consuming ways: John Houston of Atlanta University, Roosevelt Gentry and Corlis Powell Johnson, both of Jackson State University, and Lee Lorch of York University in Toronto, Canada. My deepest appreciation, however, is extended to the women about whom this article is written; their generous sharing of time, confidences, and encouragement have made the article possible. Talking with them has been a great privilege that has enriched my life.

References

1. Vivienne Malone Mayes, Black and female, Association for Women in Mathematics Newsletter, 5, no. 6 (1975) 4–6.
2. Black women in mathematics, tapes from the AWM panel of January 7, 1978, Atlanta, Georgia, published in the AWM Newsletter, 8, no. 3 (1978), and 10, no. 3 (1980).
3. The Double Bind: The Price of Being a Minority Woman in Science, AAAS Report No. 76-R-3, April, 1976.
4. Judy Green, American women in mathematics—The first Ph.D.'s, AWM Newsletter, 8, no. 1 (1978) 13–15.
5. Newell, Gipson, Rich, Stubblefield, Black Mathematicians and Their Work, Dorrance, Ardmore, Pa., 1980.
6. Ethel Turner, A Negro History Compendium, Cheyney State College, Cheyney, Pa., 1971.
7. Patricia Kenschaft, Marjorie Lee Browne: In memoriam, AWM Newsletter, 10, no. 5 (1980) 8–11.
8. Vivienne Malone Mayes, Lee Lorch at Fisk: A tribute, this MONTHLY, 83 (1976) 708–711.
9. Personal conversations and correspondence with the women discussed in this article.

MISCELLANEA

62. Sometimes his mind went blank. He did not like calculus; it seemed to him like a dream; there was nothing behind it, its formulations. Other textbooks did not trouble him like this—he could memorize anything that had a reality behind it. He did not have to touch this reality, to move his fingers across it, he had only to know that it existed somewhere and might be measured, might be cut out and held up triumphantly to the light.... But mathematics disturbed him. A stunning whirl of numbers, insubstantial numbers, signs with nothing behind them that somehow corresponded to ideas in the brain.... No, he could not understand.—Reprinted from *Wonderland*, by Joyce Carol Oates, by permission of the publisher, Vanguard Press, Inc. Copyright © 1971 by Joyce Carol Oates. (Suggested by Harvey Diamond.)

THE WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

L. F. KLOSINSKI
Department of Mathematics, University of Santa Clara, CA 95053

G. L. ALEXANDERSON
Department of Mathematics, University of Santa Clara, CA 95053

A. P. HILLMAN
Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131

The following results of the forty-first William Lowell Putnam Mathematical Competition, held on December 6, 1980, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, five thousand dollars, was awarded to the Department of Mathematics of **Washington University**, St. Louis, Missouri. The members of its winning team were: Kevin P. Keating, Nathan E. Schroeder, and Edward A. Shpiz; each was awarded a prize of two hundred fifty dollars.

The second prize, two thousand five hundred dollars, was awarded to the Department of Mathematics of **Harvard University**, Cambridge, Massachusetts. The members of its team were: Michael Raship, Ehud B. Reiter, and Brian F. Sheppard; each was awarded a prize of two hundred dollars.

The third prize, one thousand five hundred dollars, was awarded to the Department of Mathematics of the **University of Maryland**, College Park, Maryland. The members of its team were: Ravi B. Boppanna, Brian R. Hunt, and Eric I. Kuritzky; each was awarded a prize of one hundred fifty dollars.

The fourth prize, one thousand dollars, was awarded to the Department of Mathematics of the **University of Chicago**, Chicago, Illinois. The members of its team were Daniel J. Goldstein, Nicholas F. Reingold, and Michael P. Spertus; each was awarded a prize of one hundred dollars.

The fifth prize, five hundred dollars, was awarded to the Department of Mathematics of the **University of California**, Berkeley, California. The members of its team were Randall L. Dougherty, Lin Goldstein, and Robin A. Pemantle; each was awarded a prize of fifty dollars.

The five highest-ranking individual contestants, in alphabetical order, were **Eric D. Carlson**, Michigan State University; **Randall L. Dougherty**, University of California, Berkeley; **Daniel J. Goldstein**, University of Chicago; **Laurence E. Penn**, Harvard University; and **Michael Raship**, Harvard University. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of five hundred dollars by the Putnam Prize Fund.

The next five highest-ranking individuals, in alphabetical order, were *Joel Friedman*, Harvard University; *Fred W. Helenius*, Massachusetts Institute of Technology; *Irwin L. Jungreis*, Cornell University; *Michael J. Larsen*, Harvard University; and *Arthur S. Parker*, University of Kansas. Each of these students was awarded a prize of two hundred fifty dollars.

The following teams, named in alphabetical order, received honorable mention: *California Institute of Technology*, with team members R. Sekhar Chivukula, Peter W. Shor, John R. Stembridge; *Case Western Reserve University*, with team members Edward J. Branagan, Jr., Scott R. Fluhrer, David A. Natwick; *Massachusetts Institute of Technology*, with team members Andrew J. Bernoff, Josh D. Cohen, Lorenzo A. Sadun; *Michigan State University*, with team members Eric D. Carlson, Karl A. Dahlke, Lloyd A. Rawley; and *Princeton University*, with team members Mark P. Kleiman, Jacob Nemchyonok, and Charles H. Walter.

Honorable mention was achieved by the following thirty-three individuals, named in alphabeti-

cal order: *Michael H. Albert*, University of Waterloo; *Richard J. Beigel*, Stanford University; *David D. Chambliss*, Princeton University; *Stephen J. Curran*, Beloit College; *Marc A. Drexler*, Johns Hopkins University; *Paul Feit*, Harvard University; *Paul N. Feldman*, Yale University; *Scott R. Fluhrer*, Case Western Reserve University; *Brian R. Hunt*, University of Maryland, College Park; *Howard J. Karloff*, University of Pennsylvania; *Kevin P. Keating*, Washington University, St. Louis; *Gary R. Lawlor*, Brigham Young University; *Franklin M. Maley*, Amherst College; *Michael P. Mattis*, Harvard University; *Victor J. Milenkovic*, Harvard University; *David J. Montana*, Harvard University; *Evan Morton*, Harvard University; *Robin A. Pemantle*, University of California, Berkeley; *Matthew J. Raw*, Washington University, St. Louis; *Zinory Reichstein*, California Institute of Technology; *Ehud B. Reiter*, Harvard University; *Subir Sachdev*, Massachusetts Institute of Technology; *Lorenzo A. Sadun*, Massachusetts Institute of Technology; *Erin J. Schram*, Michigan State University; *Nathan E. Schroeder*, Washington University, St. Louis; *Peter W. Shor*, California Institute of Technology; *Edward A. Shpiz*, Washington University, St. Louis; *Michael Spertus*, University of Chicago; *James Van Buskirk*, University of Minnesota, Duluth; *Charles H. Walter*, Princeton University; *Lawrence B. Weinstein*, Yale University; *David A. Williams*, Washington University, St. Louis; *Robert L. Zako*, Harvard University.

The other individuals who achieved ranks among the top 100, in alphabetical order of their schools, were: University of Alberta, *Robert P. Morewood, Peter B. Weichman*; Amherst College, *George G. Watson*; Brown University, *Bruce A. Hendrickson*; California Institute of Technology, *R. Sekhar Chivukula, Lance J. Dixon, Scott R. Johnson, Christopher P. Lutz, David J. Muraki, John R. Stembridge*; University of California, Davis, *Theodore W. Jones, Carl A. Lundgren*; University of California, Santa Barbara, *John R. Rose, Dean Connable Wills*; University of California, Santa Cruz, *Stephen P. Carrier*; Case Western Reserve University, *David A. Natwick*; University of Chicago, *Nicholas F. Reingold*; George Mason University, *Stephen Billups*; Harvard University, *Anthony R. Barker, Michael V. Finn, James G. Propp, Jonathan S. Roberts, Brian F. Sheppard, Carlos T. Simpson, William A. Titus, Ron K. Unz, Robert J. Waldmann*; University of Illinois, Urbana-Champaign, *Jerome V. Walsh*; Iowa State University, *William R. Somsky*; Université Laval, *Pierre Tremblay*; University of Maryland, College Park, *Ravi B. Boppana, Eric I. Kuritzky*; Massachusetts Institute of Technology, *Josh D. Cohen, David Seibert*; Michigan State University, *Karl Dahlke*; University of Minnesota, Minneapolis, *Peter M. Thompson*; University of North Carolina, Chapel Hill, *Edward J. Rak*; University of Oklahoma, *Gary D. Köhler*; Oregon State University, *Gregory L. Larson*; University of Pittsburgh, *Randall S. Henry*; Pomona College, *William M. McGovern*; Princeton University, *David R. Grant, Mark P. Kleiman, James L. McInnes, David P. Roberts, Bruce K. Smith, Stephen A. Vavasis*; Rensselaer Polytechnic Institute, *David R. Iny, Gregory F. Taylor*; Rose-Hulman Institute of Technology, *Michael L. Call*; Stanford University, *Thomas C. Hales*; University of Virginia, *Mark G. Pleszkoch*; Washington University, St. Louis, *Bard Bloom, Chim-Chung Chan, Karl F. Narveson*; University of Waterloo, *Guy W. Hulbert, Duncan J. Murdoch*; Yale University, *Alan S. Edelman*.

There were 2043 individual contestants from 335 colleges and universities in Canada and the United States in the competition of December 6, 1980. Teams were entered by 237 institutions.

The Questions Committee of the forty-first competition consisted of E. J. Barbeau (Chairman), Joel Spencer, and K. B. Stolarsky; they proposed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1

Let b and c be fixed real numbers and let the ten points $(j, y_j), j = 1,2,\ldots,10$, lie on the parabola $y = x^2 + bx + c$. For $j = 1,2,\ldots,9$, let I_j be the point of intersection of the tangents to the given parabola at (j, y_j) and

$(j+1, y_{j+1})$. Determine the polynomial function $y = g(x)$ of least degree whose graph passes through all nine points I_j.

Problem A-2

Let r and s be positive integers. Derive a formula for the number of ordered quadruples (a, b, c, d) of positive integers such that

$$3^r \cdot 7^s = \operatorname{lcm}[a, b, c] = \operatorname{lcm}[a, b, d] = \operatorname{lcm}[a, c, d] = \operatorname{lcm}[b, c, d].$$

The answer should be a function of r and s.

(Note that 1 cm $[x, y, z]$ denotes the least common multiple of x, y, z.)

Problem A-3

Evaluate

$$\int_0^{\pi/2} \frac{dx}{1 + (\tan x)^{\sqrt{2}}}.$$

Problem A-4

(a) Prove that there exist integers a, b, c, not all zero and each of absolute value less than one million, such that

$$|a + b\sqrt{2} + c\sqrt{3}| < 10^{-11}.$$

(b) Let a, b, c be integers, not all zero and each of absolute value less than one million. Prove that

$$|a + b\sqrt{2} + c\sqrt{3}| > 10^{-21}.$$

Problem A-5

Let $P(t)$ be a nonconstant polynomial with real coefficients. Prove that the system of simultaneous equations

$$0 = \int_0^x P(t)\sin t\, dt = \int_0^x P(t)\cos t\, dt$$

has only finitely many real solutions x.

Problem A-6

Let C be the class of all real valued continuously differentiable functions f on the interval $0 \leqslant x \leqslant 1$ with $f(0) = 0$ and $f(1) = 1$. Determine the largest real number u such that

$$u \leqslant \int_0^1 |f'(x) - f(x)|\, dx$$

for all f in C.

Problem B-1

For which real numbers c is $(e^x + e^{-x})/2 \leqslant e^{cx^2}$ for all real x?

Problem B-2

Let S be the solid in three-dimensional space consisting of all points (x, y, z) satisfying the following system of six simultaneous conditions:

$$x \geqslant 0, \quad y \geqslant 0, \quad z \geqslant 0,$$

$$x + y + z \leqslant 11,$$

$$2x + 4y + 3z \leqslant 36,$$

$$2x + 3z \leqslant 24.$$

(a) Determine the number v of vertices of S.

(b) Determine the number e of edges of S.

(c) Sketch in the bc-plane the set of points (b, c) such that $(2,5,4)$ is one of the points (x, y, z) at which the linear function $bx + cy + z$ assumes its maximum value on S.

Problem B-3

For which real numbers a does the sequence defined by the initial condition $u_0 = a$ and the recursion $u_{n+1} = 2u_n - n^2$ have $u_n > 0$ for all $n \geqslant 0$?

(Express the answer in the simplest form.)

Problem B-4

Let $A_1, A_2, \ldots, A_{1066}$ be subsets of a finite set X such that $|A_i| > \frac{1}{2}|X|$ for $1 \leqslant i \leqslant 1066$. Prove there exist ten elements $x_1, \ldots, x_{10}$ of X such that every A_i contains at least one of $x_1, \ldots, x_{10}$.

(Here $|S|$ means the number of elements in the set S.)

Problem B-5

For each $t \geqslant 0$, let S_t be the set of all nonnegative, increasing, convex, continuous, real-valued functions $f(x)$ defined on the closed interval $[0, 1]$ for which

$$f(1) - 2f(2/3) + f(1/3) \geqslant t[f(2/3) - 2f(1/3) + f(0)].$$

Develop necessary and sufficient conditions on t for S_t to be closed under multiplication.

(This closure means that, if the functions $f(x)$ and $g(x)$ are in S_t, so is their product $f(x)g(x)$. A function $f(x)$ is convex if and only if $f(su + (1 - s)v) \leqslant sf(u) + (1 - s)f(v)$ whenever $0 \leqslant s \leqslant 1$.)

Problem B-6

An infinite array of rational numbers $G(d, n)$ is defined for integers d and n with $1 \leqslant d \leqslant n$ as follows:

$$G(1, n) = \frac{1}{n}, \qquad G(d, n) = \frac{d}{n}\sum_{i=d}^{n} G(d-1, i-1) \quad \text{for} \quad d > 1.$$

For $1 < d \leqslant p$ and p *prime*, prove that $G(d, p)$ is expressible as a quotient s/t of integers s and t with t *not* an integral multiple of p.

(For example, $G(3,5) = 7/4$ with the denominator 4 not a multiple of 5.)

On the line of the problem number for each solution, a 12-tuple $(n_{10}, n_9, \ldots, n_0, n_{-1})$ is given in which the entry n_i for $10 \geqslant i \geqslant 0$ is the number of contestants among the top 207 who achieved i points for the problem, and n_{-1} is the number not submitting a solution.

SOLUTIONS

A-1. (52, 106, 7, 9, 0, 0, 0, 1, 8, 8, 8, 8)

We show that $g(x) = x^2 + bx + c - (1/4)$. The equation of the tangent to the given parabola at $P_j = (j, y_j)$ is easily seen to be $y = L_j$, where $L_j = (2j + b)x - j^2 + c$. Solving $y = L_j$ and $y = L_{j+1}$ simultaneously, one finds that $x = (2j + 1)/2$ and so $j = (2x - 1)/2$ at I_j. Substituting this expression for j into L_j gives the $g(x)$ above.

A-2. (142, 0, 0, 0, 0, 0, 0, 0, 14, 27, 13, 11)

We show that the number is $(1 + 4r + 6r^2)(1 + 4s + 6s^2)$. Each of a, b, c, d must be of the form $3^m 7^n$ with m in $\{0, 1, \ldots, r\}$ and n in $\{0, 1, \ldots, s\}$. Also m must be r for at least two of the four numbers, and n must be s for at least two of the four numbers. There is one way to have $m = r$ for all four numbers, $4r$ ways to have one m in $\{0, 1, \ldots, r - 1\}$ and the other three equal to r, and $\binom{4}{2}r^2 = 6r^2$ ways to have two of the m's in $\{0, 1, \ldots, r - 1\}$ and the other two equal to

r. Thus there are $1 + 4r + 6r^2$ choices of allowable m's and, similarly, $1 + 4s + 6s^2$ choices of allowable n's.

A-3. (20, 2, 1, 0, 0, 0, 0, 0, 2, 2, 39, 141)

Let I be the given definite integral and $\sqrt{2} = r$. We show that $I = \pi/4$. Using $x = (\pi/2) - u$, one has

$$I = \int_{\pi/2}^{0} \frac{-du}{1 + \cot^r u} = \int_0^{\pi/2} \frac{\tan^r u \, du}{\tan^r u + 1}.$$

Hence

$$2I = \int_0^{\pi/2} \frac{1 + \tan^r x}{1 + \tan^r x} dx = \int_0^{\pi/2} dx = \pi/2 \quad \text{and} \quad I = \pi/4.$$

A-4. (4, 6, 2, 0, 0, 0, 0, 1, 23, 10, 29, 132)

(a). Let S be the set of the 10^{18} real numbers $r + s\sqrt{2} + t\sqrt{3}$ with each of r, s, t in $\{0, 1, \dots, 10^6 - 1\}$ and let $d = (1 + \sqrt{2} + \sqrt{3})10^6$. Then each x in S is in the interval $0 \leqslant x < d$. This interval is partitioned into $10^{18} - 1$ "small" intervals $(k - 1)e \leqslant x < ke$ with $e = d/(10^{18} - 1)$ and k taking on the values $1, 2, \dots, 10^{18} - 1$. By the pigeonhole principle, two of the 10^{18} numbers of S must be in the same small interval and their difference $a + b\sqrt{2} + c\sqrt{3}$ gives the desired a, b, c since $c < 10^{-11}$.

(b) Let $F_1 = a + b\sqrt{2} + c\sqrt{3}$ and F_2, F_3, F_4 be the other numbers of the form $a \pm b\sqrt{2} \pm c\sqrt{3}$. Using the irrationality of $\sqrt{2}$ and $\sqrt{3}$ and the fact that a, b, c are not all zero, one easily shows that no F_i is zero. (The demonstration of this was Problem A-1 of the 15th Competition, held on March 5, 1955. For the proof, see page 402 of *The William Lowell Putnam Mathematical Competition Problems and Solutions*: *1938–1964*, published by the MAA, or see this MONTHLY, 62 (1955) 561.) One also sees readily that the product $P = F_1F_2F_3F_4$ is an integer. Hence $|P| \geqslant 1$. Then $|F_1| \geqslant 1/|F_2F_3F_4| > 10^{-21}$ since $|F_i| < 10^7$ and thus $1/|F_i| > 10^{-7}$ for each i.

A-5. (8, 6, 6, 4, 3, 3, 0, 2, 2, 11, 55, 107)

Let $Q = P - P'' + P^{\text{iv}} - \cdots$. Using repeated integrations by parts, the equations of the given system become

$$\int_0^x P(t)\sin t \, dt = -Q(x)\cos x + Q'(x)\sin x + Q(0) = 0,$$

$$\int_0^x P(t)\cos t \, dt = Q(x)\sin x + Q'(x)\cos x - Q'(0) = 0.$$

These imply that

$$Q(x) = Q'(0)\sin x + Q(0)\cos x. \qquad (E)$$

Since P' and, hence, Q are polynomials of positive degree and the right side of (E) is bounded, equation (E) has all of its solutions in some interval $|x| \leqslant M$. In such an interval, $P(x)\sin x$ has only finitely many zeros and $\int_0^x P(t)\sin t \, dt = 0$ has at most one more zero by Rolle's Theorem. Q.E.D.

A-6. (1, 0, 1, 0, 0, 10, 0, 0, 3, 3, 73, 116)

We show that $u = 1/e$. Since $f' - f = (fe^{-x})'e^x$ and $e^x \geqslant 1$ for $x \geqslant 0$,

$$\int_0^1 |f' - f| \, dx = \int_0^1 |(fe^{-x})'e^x| \, dx \geqslant \int_0^1 (fe^{-x})' \, dx = [fe^{-x}]_0^1 = 1/e.$$

To see that $1/e$ is the largest lower bound, we use functions $f_a(x)$ defined by

$$f_a(x) = (e^{a-1}/a)x \quad \text{for} \quad 0 \leqslant x \leqslant a, \qquad f_a(x) = e^{x-1} \quad \text{for} \quad a \leqslant x \leqslant 1.$$

Let $m = e^{a-1}/a$. Then

$$\int_0^1 |f_a'(x) - f_a(x)|\,dx = \int_0^a |m - mx|\,dx = m\left(a - \frac{a^2}{2}\right) = e^{a-1}\left(1 - \frac{a}{2}\right).$$

As $a \to 0$, this expression approaches $1/e$. The function $f_a(x)$ does not have a continuous derivative, but one can smooth out the corner, keeping the change in the integral as small as one wishes, and thus show that no number greater than $1/e$ can be an upper bound.

B-1. (27, 19, 20, 0, 0, 0, 0, 41, 15, 20, 34, 31)

The inequality holds if and only if $c \geqslant 1/2$. For $c \geqslant 1/2$,

$$\frac{e^x + e^{-x}}{2} = \sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!} \leqslant \sum_{n=0}^{\infty} \frac{x^{2n}}{2^n n!} = e^{x^2/2} \leqslant e^{cx^2}$$

for all x since $(2n)! \geqslant 2^n n!$ for $n = 0, 1, \ldots$.

Conversely, if the inequality holds for all x, then

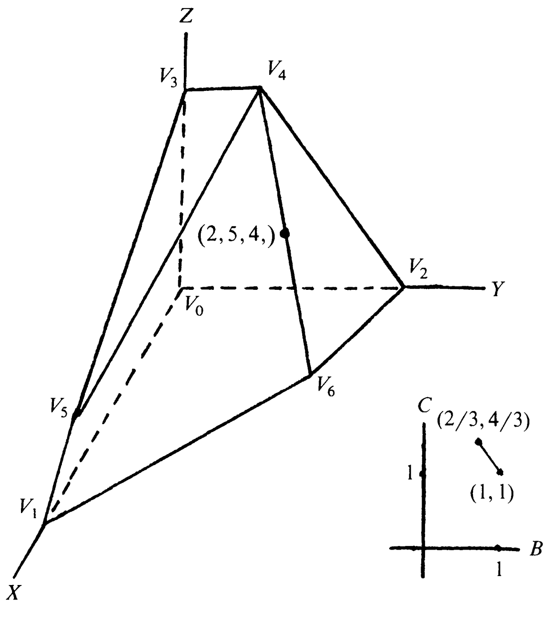
$$0 \leqslant \lim_{x\to 0} \frac{e^{cx^2} - \frac{1}{2}(e^x + e^{-x})}{x^2} = \lim_{x \to 0} \frac{(1 + cx^2 + \cdots) - \left(1 + \frac{1}{2}x^2 + \cdots\right)}{x^2} = c - \frac{1}{2}$$

and so $c \geqslant 1/2$.

B-2. (19, 4, 11, 4, 2, 0, 3, 8, 10, 41, 53, 52)

(a) $v = 7$. The seven vertices are $V_0 = (0,0,0)$, $V_1 = (11,0,0)$, $V_2 = (0,9,0)$, $V_3 = (0,0,8)$, $V_4 = (0,3,8)$, $V_5 = (9,0,2)$, and $V_6 = (4,7,0)$.

(b) $e = 11$. The eleven edges are V_0V_1, V_0V_2, V_0V_3, V_1V_5, V_1V_6, V_2V_4, V_2V_6, V_3V_4, V_3V_5, V_4V_5, and V_4V_6.



(c) The desired (b, c) are those with $b + c = 2$ and $2/3 \leqslant b \leqslant 1$. Let $L(x, y, z) = bx + cy + z$. Since L is linear and $(2,5,4)$ is on edge V_4V_6, the maximum of L on S must be assumed at V_4 and at V_6 and the conditions on b and c are obtained from $L(0,3,8) = L(4,7,0) \geqslant L(x, y, z)$, with (x, y, z) ranging over the other five vertices.

B-3. (62, 27, 29, 10, 2, 6, 5, 6, 16, 27, 2, 15)

We show that $u_n > 0$ for all $n \geqslant 0$ if and only if $a \geqslant 3$. Let $\Delta u_n = u_{n+1} - u_n$. Then the recursion (i.e., difference equation) takes the form $(1 - \Delta)u_n = n^2$. Since n^2 is a polynomial, a particular solution is

$$u_n = (1 - \Delta)^{-1}n^2 = (1 + \Delta + \Delta^2 + \cdots)n^2 = n^2 + (2n + 1) + 2 = n^2 + 2n + 3.$$

(This is easily verified by substitution.) The complete solution is $u_n = n^2 + 2n + 3 + k\cdot 2^n$, since $v_n = k\cdot 2^n$ is the solution of the associated homogeneous difference equation $v_{n+1} - 2v_n = 0$. The desired solution with $u_0 = a$ is $u_n = n^2 + 2n + 3 + (a - 3)2^n$. Since $\lim_{n\to\infty}[2^n/(n^2 + 2n + 3)] = +\infty$, u_n will be negative for large enough n if $a - 3 < 0$. Conversely, if $a - 3 \geqslant 0$, it is clear that each $u_n > 0$.

Alternatively, one sees that $u_0 = a$ and $u_1 = 2a$ and one can prove by mathematical induction that

$$u_n = 2^n a - \sum_{k=1}^{n-1} 2^{n-1-k}k^2 \quad \text{for} \quad n \geqslant 2.$$

Hence $u_n > 0$ for $n \geqslant 0$ if and only if $a > \Sigma_{k=1}^{n-1}2^{-1-k}k^2$ and this holds if and only if $a \geqslant L$, where $L = \Sigma_{k=1}^{\infty}2^{-1-k}k^2$. Let D mean d/dx. Then for $|x| < 1$,

$$(1 - x)^{-1} = \sum_{k=0}^{\infty} x^k$$

$$D(1 - x)^{-1} = (1 - x)^{-2} = \sum_{k=1}^{\infty} kx^{k-1}$$

$$D(1 - x)^{-2} = 2(1 - x)^{-3} = \sum_{k=2}^{\infty} k(k - 1)x^{k-2}.$$

Let $g(x) = 2x^3(1 - x)^{-3} + x^2(1 - x)^{-2}$. Then $L = g(1/2) = 3$ and the answer is all $a \geqslant 3$.

B-4. (1, 38, 14, 11, 1, 0, 0, 0, 2, 4, 1, 40, 95)

The result we are asked to prove is clearly not true if $|X| < 10$. Hence we assume that $|X| \geqslant 10$ or that the A_j are distinct, which implies that $|X| \geqslant 10$.

Let $X = \{x_1, \ldots, x_m\}$, with $m = |X|$, and let n_i be the number of j such that x_i is in A_j. Let N be the number of ordered pairs (i, j) such that x_i is in A_j. Then

$$N = n_1 + n_2 + \cdots + n_m = |A_1| + |A_2| + \cdots + |A_{1066}| > 1066(m/2) = 533m.$$

Hence one of the n_i, say n_1, exceeds 533.

Let $B_1, \ldots, B_s$ be those sets A_j not containing x_1 and $Y = \{x_2, x_3, \ldots, x_m\}$. Then $s = 1066 - n_1 \leqslant 532$ and each $|B_j| > |Y|/2$. We can assume that x_2 is in at least as many B_j as any other x_i and let $C_1, \ldots, C_t$ be the B_j not containing x_2. As before, one can show that $t \leqslant 265$.

We continue in this way. The 4th sequence of sets $D_1, \ldots, D_u$ will number no more than 132. The numbers of sets in the 5th through 10th sequences will number no more than 65, 32, 15, 7, 3, and 1, respectively. Thus we obtain the desired elements $x_1, \ldots, x_{10}$ unless X has fewer than 10 elements.

B-5. (0, 1, 0, 0, 0, 0, 0, 0, 1, 8, 31, 166)

The answer is $1 \geqslant t$ (or $0 \leqslant t \leqslant 1$). The product fg of two nonnegative increasing continuous

real-valued functions has the same properties. Using the fact that $0 \leqslant a \leqslant c$ and $0 \leqslant b \leqslant d$ imply $ad + bc \leqslant cb + cd$, one shows that fg is convex when f and g are convex. The function $f(x) = x$ is in S_t for all t. If S_t is closed under multiplication, x^2 is in S_t and so $2/9 = 1 - 2(4/9) + (1/9) \geqslant t[4/9 - 2(1/9)] = 2t/9$ or $1 \geqslant t$. Conversely, when $1 \geqslant t$, a lengthy, straightforward computation verifies that S_t is closed.

B-6. $(1,0,0,0,0,0,0,0,1,17,20,168)$

Let $F_d(x) = \Sigma_{n=d}^{\infty} G(d, n)x^n$. Then $F_1(x) = \Sigma_{n=1}^{\infty} x^n/n$ and $F_1'(x) = \Sigma_{n=0}^{\infty} x^n$. One sees that $F_d'(x) = dF_{d-1}(x)F_1'(x)$ by finding the coefficients of x^{n-1} on both sides and using $nG(d, n) = d\Sigma_{i=d}^{n} G(d-1, i-1)$. Then an induction gives us $F_d(x) = [F_1(x)]^d$. Now, for $1 < d \leqslant p$, the coefficient $G(d, p)$ of x^p in $F_d(x)$ is the coefficient of x^p in $[\Sigma_{n=1}^{p-d+1} x^n/n]^d$, and hence $G(d, p) = s/t$ with s and t integers and t a product of primes less than p.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

ON McCARTY'S QUEEN SQUARES

CYRIL W. L. GARNER
Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6

AGNES M. HERZBERG
Department of Mathematics, Imperial College, London, England

1. Introduction. McCarty [**3**] introduced queen squares and posed questions regarding their existence. These squares arise as an extension of the well-known problem of placing n sets of n nonattacking queens on an $n \times n$ chess-board; see, for example, Rouse Ball and Coxeter [**1**].

According to McCarty, a *queen square* of order n is a square arrangement of the elements from $\{1, 2, \ldots, n\}$ in an $n \times n$ array such that each element occurs at most once in each row, column, and diagonal, and elements i and j may not be placed $|i - j|$ entries apart in any row, column, or diagonal. Note that there may be empty grid squares.

McCarty represented the maximum number of nonempty entries that a queen square of order n contains by $M(n)$ and denoted by $R(n) = n^{-2}M(n)$ the ratio of $M(n)$ to the number of grid squares of an $n \times n$ board. He gave a table of values of lower bounds of $M(n)$ and $R(n)$ for from 3 through 18.

In a recent paper [**2**], the authors defined Latin queen squares. These are the same as queen squares except that no empty grid squares are allowed in the $n \times n$ array. They showed that Latin queen squares exist for every prime number p greater than or equal to 11.

The purpose of this brief note is to exhibit the relationship between queen squares and Latin queen squares and to show that, for n a prime greater than or equal to 11, $M(n) = n^2$ and $R(n) = 1$. This answers the first of the following questions posed by McCarty:

(1) Can a latin square be constructed that is also a queen square? If yes, for which n is it possible? If no, is there an upper bound on $R(n)$?

(2) Does there exist an algorithm for the maximal placement of the queens into the n-cube?

See [2] for an answer to (2) when n is a prime greater than or equal to 11.

2. Examples of Latin Queen Squares. The authors have given a constructive proof of the existence of many Latin queen squares for all primes greater than or equal to 11, thus exhibiting queen squares where all entries are filled. One such construction is the following:

1	3	5	7	9	$\cdots$		$p-4$	$p-2$	p	2	4	6	$\cdots$	$p-5$	$p-3$	$p-1$
5	7	9			$\cdots$	$p-2$	p	2	4	6			$\cdots$	$p-1$	1	3
9	11				$\cdots$	2	4	6	8	10			$\cdots$	3	5	7
$\cdots$																
$p-7$	$p-5$	$p-3$	$p-1$	1	3			$\cdots$								$p-9$
$p-3$	$p-1$	1	3	5	7			$\cdots$								$p-5$.

Thus, for $p = 11$,

1	3	5	7	9	11	2	4	6	8	10
5	7	9	11	2	4	6	8	10	1	3
9	11	2	4	6	8	10	1	3	5	7
2	4	6	8	10	1	3	5	7	9	11
6	8	10	1	3	5	7	9	11	2	4
10	1	3	5	7	9	11	2	4	6	8
3	5	7	9	11	2	4	6	8	10	1
7	9	11	2	4	6	8	10	1	3	5
11	2	4	6	8	10	1	3	5	7	9
4	6	8	10	1	3	5	7	9	11	2
8	10	1	3	5	7	9	11	2	4	6,

and, for $p = 13$,

1	3	5	7	9	11	13	2	4	6	8	10	12
5	7	9	11	13	2	4	6	8	10	12	1	3
9	11	13	2	4	6	8	10	12	1	3	5	7
13	2	4	6	8	10	12	1	3	5	7	9	11
4	6	8	10	12	1	3	5	7	9	11	13	2
8	10	12	1	3	5	7	9	11	13	2	4	6
12	1	3	5	7	9	11	13	2	4	6	8	10
3	5	7	9	11	13	2	4	6	8	10	12	1
7	9	11	13	2	4	6	8	10	12	1	3	5
11	13	2	4	6	8	10	12	1	3	5	7	9
2	4	6	8	10	12	1	3	5	7	9	11	13
6	8	10	12	1	3	5	7	9	11	13	2	4
10	12	1	3	5	7	9	11	13	2	4	6	8

are Latin queen squares.

References

1. W. W. Rouse Ball and H. S. M. Coxeter, Mathematical Recreations and Essays, 12th ed., University of Toronto Press, 1974, pp. 165–173.
2. A. M. Herzberg and C. W. L. Garner, Latin queen squares (to appear).
3. C. P. McCarty, Queen squares, this MONTHLY, 85 (1978) 578–580.

ADDENDUM TO "HOW TO CUT A CAKE FAIRLY"

[This MONTHLY, 87 (1980) 640–644]

WALTER STROMQUIST
Daniel H. Wagner Associates, Station Square One, Paoli, PA 19301

"How to Cut a Cake Fairly," by Walter Stromquist, appeared in the October 1980 Mathemati-

cal Notes Section of this MONTHLY. The author wishes to call attention to a paper by D. R. Woodall [1] which appeared at about the same time and contains the same result (with a different and independent proof). Woodall's paper includes a description of J. L. Selfridge's elegant algorithm for dividing a cake among three people. Another paper on the subject of cake cutting, by Rebman [2], also deserves mention.

One passage in "How to Cut a Cake Fairly" requires correction. The third sentence of the article was intended to read, correctly,

> In a simpler version of the problem, a division is regarded as "fair" if each person ("player") is satisfied that he has received at least $1/n$ of the cake.

In the MONTHLY, the sentence was changed to conclude

> ... if all people ("players") are satisfied that each has received at least $1/n$ of the cake.

The change was made, along with a few others, in the final proof, without informing the author. The MONTHLY regrets the unintentional change in meaning.

References

1. D. R. Woodall, Dividing a cake fairly, J. Math. Analysis and Applications, 78 (1980) 233–247.
2. K. Rebman, How to get (at least) a fair share of the cake, in Mathematical Plums, Dolciani Mathematical Expositions No. 4, ed. by R. Honsberger, M.A.A., 1979, pp. 22–37.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

A GENERALIZATION OF THE FORMULA FOR COMPUTING THE INVERSE OF A MATRIX

CHANTAL SHAFROTH
Mathematics Department, North Carolina Central University, Durham, NC 27707

A classical method for finding the inverse of a nonsingular square matrix M, element by element, consists of using the following formula:

$$x_{ij} = \frac{(-1)^{i+j} A_{ji}}{\det M} \tag{1}$$

where $M^{-1} = (x_{ij})$ and A_{ji} is the minor of the element a_{ji} of M.

Let M be of order n. Since x_{ij} is a determinant of order 1 of M^{-1}, while A_{ji} is a determinant of order $(n-1)$ of M, (1) suggests a more general relationship between the determinant of any submatrix of order m of M and the determinant of a certain submatrix of order $(n-m)$ of M^{-1}.

LEMMA. *Let M be a nonsingular square matrix,*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

whose inverse is

$$M^{-1} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix},$$

where A' and D' are square matrices of the same order as A and D, respectively; then $\det A = \det D' \cdot \det M$.

Proof. Let M be a nonsingular square matrix, which is partitioned into

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

where A and D are square matrices.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} I & 0 \\ 0 & D' \end{bmatrix} = \begin{bmatrix} A & BD' \\ C & DD' \end{bmatrix}; \tag{2}$$

I is the identity matrix of the same order as A.

The inverse of M is

$$M^{-1} = \begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix};$$

therefore

$$AB' + BD' = 0 \quad \text{and} \quad CB' + DD' = I.$$

Equation (2) becomes:

$$M \cdot \begin{bmatrix} I & 0 \\ 0 & D' \end{bmatrix} = \begin{bmatrix} A & -AB' \\ C & I - CB' \end{bmatrix}.$$

Taking the determinant of both sides we get:

$$\det M \cdot \det D' = \det \begin{bmatrix} A & -AB' \\ C & I - CB' \end{bmatrix}$$

$$= \det \begin{bmatrix} A & 0 \\ C & I \end{bmatrix} = \det A. \qquad \square$$

Notice that in the case where M is an orthogonal matrix, that is, when M^{-1} is equal to the transpose of M, the previous lemma becomes: $\det A = \det M \cdot \det D^t$, where D^t is the transpose of D, or

$$\det A = \det D, \qquad \text{if } \det M = 1,$$

and

$$\det A = -\det D, \qquad \text{if } \det M = -1.$$

The following interesting theorem was, in fact, first proved by Jacobi in 1834 (see [**1**, pp. 206–212]) but has been neglected by modern authors.

THEOREM. *Given a nonsingular square matrix M and its inverse M^{-1}, let A be the square submatrix of M obtained by deleting all rows except the* pth, qth, ... rth *rows of M and deleting all columns except the* sth, tth, ... vth *columns of M. $p < q \cdots < r$, $s < t \cdots < v$. Let D' be the square submatrix obtained by deleting the* sth, tth, ... vth *rows of M^{-1} and the* pth, qth, ... rth *columns of M^{-1}; then*

$$\det A = (-1)^{\delta} \cdot \det M \cdot \det D',$$

where

$$\delta = p + q + \cdots + r + s + t + \cdots + v.$$

Proof. By performing $(p-1)$ row exchanges, we can move the pth row of M to the first position, while the first row moves to the second position, the second one to the third, etc. Then, by performing $(q-2)$ row exchanges, we can move the qth row to the second position, while the initial first row of M moves to the third position, etc. We can perform the same exchanges for the

columns. So, if A has m rows and m columns, a sum total of Δ row and column exchanges will replace M by M_1, where

$$M_1 = \begin{bmatrix} A & \\ & D \end{bmatrix},$$

and

$$\Delta = (p-1) + (q-2) + \cdots + (r-m) + (s-1) + (t-2) + \cdots + (v-m).$$

If δ is defined as $\delta = p + q + \cdots + r + s + t + \cdots + v$, then

$$\Delta = \delta - 2(1 + 2 + \cdots + m) \equiv \delta (\text{modulo } 2).$$

Therefore, $\det M_1 = (-1)^{\delta} \cdot \det M$. The inverse, M_1^{-1}, may be obtained by performing a sum total of Δ row and column exchanges on M^{-1}, since any exchange of rows of M results in the exchange of corresponding columns of M^{-1} and vice versa. The result is:

$$M_1^{-1} = \begin{bmatrix} A' & \\ & D' \end{bmatrix},$$

where D' is the square matrix of order $(n-m)$ obtained by deleting the sth, tth,..., vth rows of M^{-1} and the pth, qth,..., rth columns of M^{-1}. Applying the previous lemma, we get:

$$\det A = \det D' \cdot \det M_1 = (-1)^{\delta} \cdot \det D' \cdot \det M. \qquad \square \quad (3)$$

Formula (1) is indeed a special case of (3).

An interesting corollary is the following:

COROLLARY. *Given that a unimodular matrix is one such that the determinants of all its submatrices are equal to* 0, +1, *or* −1, *the inverse of a nonsingular unimodular matrix is unimodular.*

Let M be a unimodular nonsingular matrix and let M^{-1} be its inverse. Let D' be any submatrix of M^{-1}; by the previous theorem:

$$\det D' = \frac{\det A}{(-1)^{\delta} \cdot \det M},$$

where A is a submatrix of M. Since $\det M \neq 0$, $|\det M| = 1$, and $\det D' = 0$, or 1 or −1; therefore M^{-1} is unimodular. $\square$

Reference

1. Thomas Muir, The History of Determinants in the Historical Order of Development, London, 1906, vol. 1, pp. 206–212.

A DIRECT DERIVATION OF A JACOBIAN IDENTITY FROM ELLIPTIC FUNCTIONS

I. J. SCHOENBERG
Mathematics Research Center, University of Wisconsin, Madison WI 53706

With $h > 0$ and $q = e^{-h}$ we have the identity

$$\prod_{n=0}^{\infty} \frac{(1+q^{2n+1}z)(1+q^{2n+1}z^{-1})}{(1-q^{2n+1}z)(1-q^{2n+1}z^{-1})} = \prod_{n=1}^{\infty} \left(\frac{1-q^{2n}}{1+q^{2n}} \right)^2 \sum_{n=-\infty}^{\infty} \frac{1}{\cosh(hn)} \cdot z^n, \quad (q < |z| < q^{-1}). \tag{1}$$

The left-hand side is meromorphic in the domain $0 < |z| < \infty$, and the right-hand side represents

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Contractible Loop

S 25 [1980, 60]. *Proposed by J. Mycielski, University of Colorado, Boulder.*

Let S be the spiral $\{(\cos t, \sin t, 1/t) : t > 0\}$. Is the loop $f(t) = (0, 1 + \cos t, 1 + \sin t)$, $-\pi \leqslant t \leqslant \pi$, contractible to a constant in the space $\mathbf{R}^3 \setminus S$?

Solution by Victor P. Snaith, University of Western Ontario, London, Ontario. We will apply the group of homotopy classes of loops in $\mathbf{R}^3 \setminus S$ with base point $x_0 = (0, 1, 0)$. For every loop l, $[l]$ denotes the class of l. Let p be a loop which follows f (the loop defined in the problem) within $[-\pi, 0]$; then it follows a circle in the plane of f which is tangent to f at x_0 and omits the first coil of S enclosed by f, within $[0, \pi/2]$; then it follows a similar circle which omits the second coil of S, within $[\pi/2, \pi/4]$, etc. Thus p consists of countably many circles in the plane of f, all tangent to f at x_0, and such that S goes once through each gap between those circles. Let q be a similar loop which, however, omits the f-like part of p. That is, q goes in time $[-\pi, 0]$ over the second circle of p, in time $[0, \pi/2]$ over the third circle, in time $[\pi/2, \pi/4]$ over the fourth circle, etc.

Thus it is evident that $[p][q]^{-1} = [f]$. On the other hand by a swing of p around S we can see that $[p] = [q]$. Hence $[f] = [\text{constant}]$. Q.E.D.

Arguing in a similar way Mycielski and Snaith together have proved that the fundamental group $\pi_1(\mathbf{R}^3 \setminus S)$ is trivial.

Also solved by the Chico Problem Group and by F. Cunningham, Jr. Their deformations resemble the wobbling motion of a coin as it spins down to rest followed by a shrinking to a point. Such a deformation has the advantage that the loop is homeomorphic to a circle at every instant except the last, but checking that the loop never intersects the spiral was a little more difficult than for the solution above.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by February 28, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2902. *Proposed by Linn Sennott, Illinois State University.*

Let $\Sigma a_n = a$ be a convergent series of nonnegative terms. Let s_n be the nth partial sum. Show that $\Sigma n a_n$ converges if and only if $\Sigma(a - s_n)$ converges.

E 2903. *Proposed by Louis W. Shapiro, Howard University.*

Consider walks in the northeast quadrant which start at $(0,0)$ and such that each step is one unit east or north and the points $(1,1), (3,3), \ldots, (2k+1, 2k+1), \ldots$ are forbidden. How many paths are there to $(2n, 2n)$?

E 2904. *Proposed by D. Hensley, Texas A & M University.*

Let $p > 2$ be prime. Call x *recoverable* if $x^{2^v} \equiv x \bmod p$ for some $v \geqslant 1$. Show that the set of recoverable x is permuted by squaring.

E 2905*. *Proposed by R. J. Stroeker, Erasmus Universiteit, Netherlands.*

Inside any triangle with vertices A, B, and C, a point P exists such that $\measuredangle PAB = \measuredangle PBC = \measuredangle PCA =: \omega$. The point P is called a Brocard point and the angle ω is called its Brocard angle.

If α, β, and γ are the angles of triangle ABC, then prove the inequalities:

$$\omega^{-1} < \alpha^{-1} + \beta^{-1} + \gamma^{-1} \leqslant \tfrac{3}{2}\omega^{-1} \quad \text{and} \quad \tfrac{3}{4}\omega^{-2} \leqslant \alpha^{-2} + \beta^{-2} + \gamma^{-2} < \omega^{-2}.$$

E 2906. *Proposed by Jack Garfunkel, Flushing, N.Y.*

Let I be the incenter of triangle ABC. A', B', C' are the intersections of AI, BI, CI with the incircle of ABC. Continue the process by defining I' (incenter of $A'B'C'$), then $A''B''C''$, etc. Prove that the angles of triangle $A^{(n)}B^{(n)}C^{(n)}$ approach $\pi/3$.

E 2907. *Proposed by H. Kestelman, University College, London.*

Given that the sequence $A, A^2, A^3, \ldots$ converges to a nonzero matrix A^∞, show that $A^\infty = V(WV)^{-1}W$ where V is any matrix whose columns are a basis of the right kernel of $A - I$, and W is any one whose rows are a basis of the left kernel of $A - I$.

SOLUTIONS OF ELEMENTARY PROBLEMS

Equation $(1 - x + xy)(1 - y + yx) = 1$ in Rings

E 2825 [1980, 220]. *Proposed by R. A. Melter, Southampton College.*

In which rings is the following proposition valid:

$$(*)\ x = y \quad \text{if and only if} \quad (1 - x + xy)(1 - y + yx) = 1?$$

Composite solution by F. S. Cater, Portland State University, Chico Problems Group (Calif.), and Young L. Park, Dhahran, Saudi Arabia. The proposition is easily checked in any Boolean ring, and also in the rings $\mathbb{Z}/(4)$, $\mathbb{Z}/(8)$. It is invalid in $\mathbb{Z}/(2)$. Let S denote the property: if $x^2 = x$ then $x = 1$. Let B denote the property $\forall_x[x^2 = x]$. Then $S \,\&\, (*) \Rightarrow B$. *Proof.* (**) $(1 - x + x^2)^2 = 1 \Rightarrow 1 - x + x^2 = 1 \Rightarrow x^2 = x$. □

If (**) holds then $9 = 1$, i.e., $8 = 0$. Not all rings satisfying (*) have necessarily been discovered.

A Degenerate Intersection in Metric Space

E 2826 [1980, 303]. *Proposed by Heinz W. Engl and Lewis Lum, University of Delaware.*

Prove or disprove: Let X be a nondegenerate uniquely arcwise connected compact metric space and $\{D_n\}_{n=1}^{\infty}$ a decreasing $(D_{n+1} \subseteq D_n)$ sequence of dense subsets with arcwise connected complements. If $\cap_{n=1}^{\infty}\operatorname{int}(D_n) \neq \varnothing$, then $\cap_{n=1}^{\infty} D_n$ is nondegenerate.

Solution by the proposers. Counterexample: For each point $(c, 0) \in C \times \{0\}$, where C is the Cantor set, let $L(c)$ denote the straight line segment from $(1/2, 1/2)$ to $(c, 0)$. Then $X = \cup\{L(c) | c \in C\}$, with the relative topology, is compact and any two points in X are joined by one and only one arc in X. Let $E = \{x_i\}_{i=1}^{\infty}$ be any enumeration of the endpoints of the deleted intervals arising in the construction of C. For each n let $D_n = (B_n \cup \cup_{i=n}^{\infty} L(x_i)) - (1/2, 1/2)$ where B_n is the open ball in X of radius $1/n$ about $(0, 0)$. Each D_n is dense, $D_{n+1} \subseteq D_n$ and $X - D_n$ is arcwise connected. But $\cap_{n=1}^{\infty} D_n = \cap_{n=1}^{\infty}\operatorname{int}(D_n) = \{(0, 0)\}$.

Also solved by F. S. Cater and M. D. Meyerson.

Finite Diameter, Infinite Length

E 2829 [1980, 303]. *Proposed by Mark Meyerson, U.S. Naval Academy, Annapolis.*

Prove or disprove: In a connected metric space of finite diameter, in which every point has a neighborhood homeomorphic to E^n for some fixed n, every pair of points can be connected by a path (a continuous map of an interval) of finite length. (The length of a path is defined as the limit

of the sum of the distances between the images of the endpoints of the subintervals in a subdivison of the interval as the norm of the subdivision approaches zero.)

Solution by John Cantwell, St. Louis University. Koch's snowflake curve (see for example figure 13.2, pg. 80, Michael Henle, *A Combinatorial Introduction to Topology*, W. H. Freeman, San Francisco, 1979) is metrizable (as a subset of the plane), homeomorphic to the circle (therefore compact, connected, and the metric is bounded), and clearly *no* pair of points can be joined by a path of finite length.

A question generalizing the above is: Does there exist a metric on the 2-sphere so that no pair of distinct points can be joined by a path of finite length?

Also solved by M. D. Ašić (Yugoslavia), F. S. Cater, Gesing Leung (Hong Kong), O. P. Lossers (Netherlands), M. J. Reed, D. M. Wells, and the proposer.

A Permutation of Pasta

E 2832 [1980, 404]. *Proposed by David Merriell, Vassar College.*

Take a dish containing n strands of cooked spaghetti where n is large enough and each strand is long enough so that it is not obvious whether two ends belong to the same strand. Reach in, select two ends at random (independently) and join them with edible paste. Continue to select two unjoined ends and join them until there are no more ends. One will then have a certain number of loops of spaghetti. What are the expected total number of loops and the expected number of 1-loops, i.e., loops of length one?

Solution by Victor Hernandez, Segovia, Spain; O. P. Lossers, Eindhoven University of Technology, Eindhoven, the Netherlands; Milton Sobel, University of California at Santa Barbara, and the St. Olaf College Problem Solving Group. The answers are

$$\sum_{k=1}^{n} \frac{1}{2k-1} \quad \text{and} \quad \frac{n}{2n-1}.$$

For the first part, let L_n denote the expected total number of loops, starting with n strands, and let P_n denote the probability that the first pair of ends joined together form a 1-loop. Computing expected values conditional on the result of the first join, we have

$$L_n = P_n(1 + L_{n-1}) + (1 - P_n)L_{n-1} = P_n + L_{n-1} = 1/(2n-1) + L_{n-1}.$$

From this and the fact that $L_1 = 1$, we get $L_n = 1 + \frac{1}{3} + \frac{1}{5} + \cdots + 1/(2n-1)$.

For the second part, label the strands $1, 2, \ldots n$; let

$$X_i = \begin{cases} 1 & \text{if } i\text{th strand forms a 1-loop} \\ 0 & \text{otherwise} \end{cases}$$

and let $X = \Sigma X_i$. The probability that $X_i = 1$ is $1/(2n-1)$ for all i and the expected number of 1-loops is

$$E(X) = \sum_{i=1}^{n} E(X_i) = nE(X_i) = nP(X_i = 1) = n \cdot \frac{1}{2n-1}.$$

Also solved by Edward Brody, Chico Problem Group, Peter de Buda (Canada), Milton Eisner, Noel Glick, David Gootkind, Elgin Johnston, W. T. M. Kars (Netherlands), J. Schaer (Canada), Allen Schwenk, Peter Shoenfield, Abraham Smuckler (Israel), David Wells, and David Weinman.

Lennard Råde points out that the first part of the problem is solved in A. Engel, *Wahrscheinlichkeitsrechnung und Statistik*, Band 2, Klett Verlag, 1976, page 91.

Several correspondents (correctly) interpreted the problem in terms of cycles in permutations but (incorrectly) assumed the permutations to be equally likely.

$\phi(n) = \phi(n-1) + \phi(n-2)$

E 2833 [1980, 404]. *Proposed by Anders Bagers, Denmark.*

Call an integer $n \geqslant 3$ a Phibonacci number if

$$\phi(n) = \phi(n-1) + \phi(n-2),$$

where ϕ denotes the Euler ϕ-function. Is there any composite Phibonacci number?

Solution. The answer is yes, 1037 being the smallest one. It was found by all the solvers mentioned in this summary. P. J. Weinberger (Bell Laboratories, Murray Hill, N.J.) found 70 Phibonacci numbers $\leqslant$ 200,000,000, of which 46 are composite. Harry Nelson (Lawrence Livermore Laboratories) kindly checked the computations. The largest prime Phibonacci number in the tabulation is 188,008,897; the largest composite one is 197,389,781. All are odd.

The possible existence of an even Phibonacci number was discussed by Thomas McConnell (Upper Montclair, N.J.), by Stuart Clary (New York, N.Y.), and by Frank Meyer & Jim Holmes (Bethel College). If it exists, it exceeds 10^{1600}.

The other solvers were W. A. Al-Salam (Canada), R. Breusch, D. Jurca, O. P. Lossers (Netherlands), P. Q. Perlmutter & S. Janke, R. F. Poppen, I. Rosenholtz, R. E. Stone (student), J. van de Lune (Netherlands), and C. Zimmerman.

D. Finkel studied the relation $\phi(n) = \phi(n-1) + \phi(n-2) + a$ for small values of a. D. E. Penney studied $\sigma(n) = \sigma(n-1) + \sigma(n-2)$, $\tau(n) = \tau(n-1) + \tau(n-2)$ (sum of divisors, number of divisors).

Weighted Sums of Sides and Medians

E 2837* [1980, 489]. *Proposed by C. W. Scherr, University of Texas at Austin.*

Let a_{ij} be the side of a triangle that connects vertices i and j. Let m_i be the median from vertex i. Elementary application of the law of cosines yields the relation.

$$a_{12}^{\alpha} + a_{23}^{\alpha} + a_{31}^{\alpha} = \lambda^{\alpha/2}(m_1^{\alpha} + m_2^{\alpha} + m_3^{\alpha}),$$

valid for all triangles when $\alpha = 2$ or $\alpha = 4$ and $\lambda = 4/3$. (i) Find an expression for λ in the limit as α goes to zero. (ii) Find the class of triangles for which the relation is valid for a fixed λ and arbitrary α.

Solution to (i) *by R. Breusch, Amherst, Mass., Paul S. Bruckman, Concord, Calif., L. Kuipers, Mollens Vs, Switzerland, C. V. Nakassis, Gaithersburg, Maryland, C. R. Pranesachar, S.D.G.S. College, Hindupur, India, O. G. Ruehr, Michigan Technological University, and E. G. Straus, University of California at Los Angeles.* The limiting value $\alpha \to 0$ gives (using $x^{\alpha} \approx 1 + \alpha \log x + O(\alpha^2)$)

$$\lambda \to (a_{12}a_{23}a_{31}/m_1m_2m_3)^{2/3}.$$

Solution to (ii) *by most of the same solvers.* If the relation

$$(3a_{12}^2)^{\beta} + (3a_{23}^2)^{\beta} + (3a_{31}^3)^{\beta} = (4m_1^2)^{\beta} + (4m_2^2)^{\beta} + (4m_3^2)^{\beta}$$

holds for all β ($\lambda = 4/3, \beta = \alpha/2$), then each term on the left is equal to a (corresponding) term on the right, possibly in permuted order. The condition for this, if $a_{12} \geqslant a_{23} \geqslant a_{31}$, is: $m_2 \geqslant m_1 \geqslant m_3$, so that $a_{12} = \lambda^{1/2}m_2$, $a_{23} = \lambda^{1/2}m_1$, $a_{31} = \lambda^{1/2}m_3$. Finally, it is true that $2a_{23}^2 = a_{12}^2 + a_{31}^2$. This is the desired condition.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA) by February 28, 1982. The solver's full post-office address should be on each sheet.

6356. *Proposed by C. R. Wall, Trident Technical College, Charleston, S.C.*

A number N is called deficient if the sum of its proper divisors is less than N: $\sigma(N) < 2N$. It is called abundant if $\sigma(N) > 2N$.

(a) Let k be fixed. Do there exist sequences of k consecutive abundant numbers?

(b) Prove that there are infinitely many sequences of 5 consecutive deficient numbers.

6357. *Proposed by Zachary Robinson, student, Massachusetts Institute of Technology.*

Let X be a nonempty set. Let $\mathbb{S}$ be a collection of topologies on X, well ordered by strict inclusion. Let τ be the topology generated by the base $\cup\mathbb{S}$. Prove or disprove: If τ is second countable, then $\mathbb{S}$ is countable.

6358. *Proposed by Simon Fitzpatrick and Lee Rubel, University of Illinois at Urbana–Champaign.*

The periodic functions on the real line $\mathbb{R}$ (of possibly different periods) form a partially ordered set under pointwise $\leqslant$. Do they form a lattice—that is, is there a sup and an inf ($\vee$ and $\wedge$) of any two periodic functions, not necessarily the pointwise sup and inf? How about the functions that are periodic and bounded? Periodic and continuous?

6359. *Proposed by Ronald J. Evans, University of California, San Diego.*

Let χ be a nontrivial Dirichlet character, and let α be complex. Show that $\Sigma\chi(p)(\log p)^{\alpha}p^{-s}$ converges for $\mathrm{Re}(s) \geqslant 1$, where the sum is over all primes p.

6360. *Proposed by Emeric Deutsch, Polytechnic Institute of New York.*

Let A be a nonnegative irreducible $n \times n$ matrix with Perron root r. Show that

$$\mathrm{tr}\Big[(rI - A)^T \mathrm{adj}\,(rI - A)\Big] \leqslant 0,$$

where I is the $n \times n$ identity matrix, adj X is the classical adjoint of X, X^T is the transpose of X, and tr X is the trace of X. Find necessary and sufficient conditions for equality to hold.

6361. *Proposed by Paul R. Chernoff, University of California, Berkeley.*

Let D be a division ring, $a \in D^*$, the group of nonzero elements. Suppose that a has only finitely many conjugates. Prove that a is in k, the center of D. (Application: a polynomial p with coefficients in k, irreducible over k but with a root in D, has infinitely many roots in D.)

SOLUTIONS OF ADVANCED PROBLEMS

Integrals of Harmonic Functions

6280 [1979, 793]. *Proposed by David Siegel, Stanford University.*

Let u be a harmonic function in a regular n-gon with sides $s_1, \ldots, s_n$ and radii $r_1, \ldots, r_n$ joining the center to the vertices. Show that

$$\sum_{i=1}^{n} \int_{s_i} u\, ds = 2\sin\frac{\pi}{n} \sum_{i=1}^{n} \int_{r_i} u\, ds,$$

where the integrals are taken with respect to arc length.

Solution by Robert B. Israel, University of British Columbia. Some assumption is needed on the relation of the boundary values of u to the function in the interior. We will assume u is continuous on the n-gon and harmonic in the interior; some weakening of this may be possible, but the statement would be wrong if we only assumed, say, that u is continuous at almost every point of the boundary.

We may approximate u uniformly by a function harmonic in a neighborhood of the n-gon (e.g., $u_r(z) = u(rz)$ for $0 < r < 1$); so it suffices to prove the formula for such functions. The real and imaginary parts of these are real parts of analytic functions in a neighborhood of the n-gon, which by Runge's Theorem can be uniformly approximated by polynomials. So it is enough to prove the formula for $u(z) = z^k$ for each nonnegative integer k. We may assume the n-gon is centered at 0 and one radius goes from 0 to 1. If k is not a multiple of n, both sides of the equation are 0 because rotation by $2\pi/n$ multiplies u by $e^{2\pi ik/n}$. If k is a multiple of n, the left-hand side is

$$2n\sin\frac{\pi}{n}\int_0^1\left(t+(1-t)e^{2\pi i/n}\right)^k dt = 2n\sin\frac{\pi}{n}(k+1)^{-1}\left(1-e^{2\pi i/n}\right)^{-1}\left(1-e^{2\pi i(k+1)/n}\right)$$
$$= 2n\sin\frac{\pi}{n}(k+1)^{-1},$$

while the right-hand side is

$$2n\sin\frac{\pi}{n}\int_0^1 t^k\,dt = 2n\sin\frac{\pi}{n}(k+1)^{-1}$$

as well.

Also solved by Ivan Netuka, Charles University, Prague, and the proposer. Netuka adds the following remark. This property of harmonic functions has been studied in a paper of G. Choquet and J. Deny (Bull. Soc. Math. France, 72 (1944) 118–140), where a reference to a relevant article of G. Bilger is also mentioned.

Star-Shaped Subsets of Banach Space

6283 [1979, 869]. *Proposed by Gordon R. Feathers, North Carolina State University at Raleigh.*

It is well known that a strongly closed convex subset of a Banach space is weakly closed. Is the same true of a strongly closed star-shaped subset?

Solution by Eero Posti, University of Joensuu, Finland. The answer is no—here is a simple counterexample:

It is easily seen that if A is star-shaped then so is $\bar{A}$. Denote by e_n $(n \geq 1)$ the nth unit-vector of l^2 and by A_n $(n \geq 2)$ the line segment from e_1 to e_n. Take $A = \cup_{n\geq 2}A_n$; then $\bar{A}$ is closed and star-shaped.

If $x \in A_n$, then there will exist λ $(0 \leq \lambda \leq 1)$ such that $x = (1-\lambda)e_1 + \lambda e_n = (1-\lambda, 0, \ldots, 0, \lambda, 0, \ldots)$. So $\|x\|^2 = (1-\lambda)^2 + \lambda^2 \geq 1/2$ implying that $0 \notin \bar{A}$. On the other hand, $e_n \to 0$ weakly. Thus 0 belongs to the weak closure of $\bar{A}$.

Also solved by Michael Humphries, Charles Riley, Thomas Starbird, and the proposer.

Sets of Primes

6285* [1980, 65]. *Proposed by Michael Brozinsky, Queensborough Community College, N.Y.*

Define the operation $*$ on the positive integers Z^+ by $a * b = ab + 2$, and call a subset S of Z^+ *anticlosed* under $*$ if and only if $a, b \in S$ implies $a * b \notin S$.

(a) Prove that there exists an infinite set T of primes such that T is either closed under $*$ or anticlosed under $*$.

(b*) Find such a set T.

Some 44 responses were received to this very easy problem. We mention here only a few remarks gleaned from these responses.

(1) No nonempty set of primes is closed. For if $p \neq 3$, then 3 divides $p * p$, while $3 * 3 = 11 \neq 3$.

(2) If P is any infinite set of positive integers, we obtain an infinite anticlosed subset T of P by

choosing t_1 in P arbitrarily and, recursively, taking t_{n+1} to be the least element of P greater than t_n^2+2.

(3) If T is the set of primes p characterized by one of the congruences $p \equiv 1 \pmod 3$, $p \equiv 2 \pmod 3$, or $p \equiv 1 \pmod 4$, then T is anticlosed. Rony Teitler (Wolfson College, Oxford) asks if the third of these sets T, augmented by 2, is a maximal anticlosed set of primes.

(4) Define $a \overset{*}{_n} b = ab + n$, $n \in Z$. Amer Bešlagić (undergraduate, University of Sarajevo, Yugoslavia) shows that no infinite set of primes is closed under $\overset{*}{_n}$. Barry Powell (Kirkland, Washington) and E. G. Straus (UCLA) observe that if m and q are such that q divides m^2+n, then the set T of primes p characterized by $p \equiv m \pmod q$ is anticlosed.

(5) Straus has suggested a less trivial version of this problem.

Can you find a sequence of primes $p_1 < p_2 < \cdots$ which is "anticlosed" under the $*$-multiplication $a * b = ab + k$ in the sense that no $*$-product of any number of primes is one of these primes?

Since $*$-multiplication is not associative we have to construct the sequence of primes carefully.

THEOREM. *There exists an infinite sequence of primes* $p_1 < p_2 < \cdots < p_n < \cdots$ *with the following properties.* (i) $p_1 > 64$; (ii) $\log\log p_n/\log p_n < 2^{-n-1}$; (iii) *no* p_n *is a star-product of a set of smaller primes.*

Proof. Assume that $p_1, \ldots, p_n$ have already been picked. We wish to consider the number of distinct star-products no greater than x made up of $m_1 + \cdots + m_n$ primes where m_i denotes the number of occurrences of p_i.

Since the star-products exceed the ordinary product we have

$$m_1 \log p_1 + \cdots + m_n \log p_n < \log x; \tag{1}$$

hence $m_i < \log x/\log p_i$.

Thus for large m_i the number of arrangements of the primes is

$$M_n = (m_1 + \cdots + m_n)!/(m_1! \cdots m_n!) < (m_1 + \cdots + m_n)^{m_1 + \cdots + m_n} m_1^{-m_1} \cdots m_n^{-m_n}$$

$$< \left[(\log p_1)^{1/\log p_1} \cdots (\log p_n)^{1/\log p_n} \left(\frac{1}{\log p_1} + \cdots + \frac{1}{\log p_n} \right)^{(1/\log p_1 + \cdots + 1/\log p_n)} \right]^{\log x} \tag{2}$$

By hypothesis (ii) we have

$$1/\log p_1 + \cdots + 1/\log p_n < 1 \quad \text{and} \quad \log\log p_1/\log p_1 + \cdots + \log\log p_n/\log p_n < \tfrac{1}{2}.$$

Hence

$$M_n < e^{(\log\log p_1/\log p_1 + \cdots + \log\log p_n/\log p_n)\log x} < x^{1/2}. \tag{3}$$

The number of associations in the star-product of m terms is

$$A_m = \frac{1}{m}\binom{2m-2}{m-1} < 4^m, \quad \text{where } m = m_1 + \cdots + m_n < \log x/\log p_1.$$

Thus the total number of associations is

$$A_m < 4^{\log x/\log p_1} < x^{1/3}. \tag{4}$$

Finally the total number of n-tuples $(m_1, \ldots, m_n)$ which satisfy (1) is less than

$$\prod_{i=1}^{n} (\log x/\log p_i + 1) < (\log x)^n \tag{5}$$

for all large x. Combining (3), (4), (5), we see that the total number of star-products less than x is

less than

$$x^{5/6}(\log x)^n < \pi(x) - \pi(x/2)$$

for large x. Hence, we can pick a prime p_{n+1} which satisfies (ii) and (iii).

Also solved by Miroslav D. Ašič (Yugoslavia), Merrill Barneby, J. G. Bliss, Ken Brown, R. C. Carson, F. S. Cater, Kathleen A. Drude, Robert L. Farrell, Barney Gallassio & Irwin Jungreis & Mike Watkins, Fred Galvin, C. T. Giel, Audrey Grabfield, Jerrold W. Grossman, Barry W. Hill-Tout, Brian R. Hunt, Yasuhiko Ikeda, Keith A. Kearnes, Pavel Kostyrko (Czechoslovakia), Detlev Laugwitz (Germany), Kin Y. Li, S. C. Locke, O. P. Lossers (Netherlands), George Mackey, Raymond Maruca, Mark D. Meyerson, Ole Miss Problem Group, Peter Schumer, Allen J. Schwenk, Eugene Spafford, E. Triesch (Germany), William C. Waterhouse, Keith Wayland (Puerto Rico), Delano P. Wegener, David G. Weinman, J. G. Wendel, Nikolay Williams, and the proposer.

Note. The proposer recognized the triviality of the problem only after it was too late to change it.

MISCELLANEA

63.

THREE HAIKU: WHAT IS MATHEMATICS?

Fire and Ice

Strange anomaly:
the flame of intuition
frozen in rigor.

Faith and Reason

Strands of axioms
intertwining with logic
in convolution.

Truth and Beauty

Crucible of proof
outshining alabaster,
outlasting marble.

KATHARINE O'BRIEN

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, P. Intersociety Preliminary Analyses on R & D in the FY 1982 Budget. AAAS, 1981, v + 142 pp, (P). Prepared by an informal consortium of 17 professional societies, this timely typescript report (published in April) summarizes research and development proposals in the entire 1982 federal budget. Includes a nine-page analysis of R & D in the mathematical and computer sciences (prepared by Truman Botts of CBMS) covering NSF, ARO, ONR, AFOSR, DOE, NIH, NASA, NBS. LAS

General, P. Transactions of the Moscow Mathematical Society, 1980, Issue 2. AMS, 1980, iv + 215 pp, $69 (P). Tom 38 (1979), comprising five papers. LAS

General, S(15-17), L. On Aesthetics in Science. Ed: Judith Wechsler. MIT Pr, 1981, 180 pp, $5.95 (P). [ISBN: 0-262-73056-1] Paperback edition of 1979 original hardcover volume (TR, May 1979). Six essays (on structured hierarchy, symmetry, quantum theory, mathematical unconscious, Darwin, rationality and intuition), introduction and bibliography from a course at M.I.T. LAS

Basic, T*(13: 1). Intermediate Algebra. Steven Kahan. HarBrace J, 1981, xii + 588 pp, $17.95. [ISBN: 0-15-541530-1] An attractive, easy-to-read treatment with a review of elementary algebra. Special features include review sections, tables summarizing information, numerous solved examples and highlighted notes anticipating common errors and offering helpful hints. JNC

Precalculus, T(13: 1). College Algebra, Eighth Edition. Paul K. Rees, Fred W. Sparks, Charles Sparks Rees. McGraw-Hill, 1981, xv + 509 pp, $17.95. [ISBN: 0-07-051733-9] Note that this is an Eighth Edition! (TR, Sixth Edition, June-July 1972; Seventh Edition, August-September 1977.) Contains essentially the same material as the Seventh Edition with 16 chapters instead of 18. Changes include combining the first three chapters into two and reducing the treatment of complex numbers from one chapter to one section. LLK

Education, P, L. Guidelines for Evaluating Computerized Instructional Materials. William P. Heck, Jerry Johnson, Robert J. Kansky. NCTM, 1981, 30 pp, $3.75 (P). [ISBN: 0-87353-176-0] A useful, brief introduction to issues that teachers inexperienced in computer software need to consider in evaluating programs and software intended for instructional applications. Includes an evaluation check list, sample documentation sheet, an "aside" to programmers of instructional material (the only part that assumes any knowledge of programming), and numerous references. Anyone with even a little experience will likely find the pamphlet too general to be useful. But those with no experience (and some trepidation) will find it invaluable as a place to begin. LAS

History, P, L.** Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905-1911). Arthur I. Miller. A-W, 1981, xxviii + 466 pp, $27.50 (P); $39.50. [ISBN: 0-201-04680-6] An historical exegesis of Einstein's seminal 1905 paper "On the electrodynamics of moving bodies," a "cornerstone of every subsequent fundamental physical theory." Miller sets forth the state of electrodynamics, optics, and electromagnetic theory at the turn of the century as historical context for Einstein's paper, analyses the internal structure of the 1905 paper, and examines subsequent developments that determined the interpretation of Einstein's work. The volume opens with a photographic "cast of characters," from Newton to Born, and concludes with Miller's translation of Einstein's paper, a version that corrects misprints in the original and in earlier translations. LAS

Foundations, T(14-16: 1), S, L. Elements of Symbolic Logic. Hans Reichenbach. Dover, 1975, xiii + 444 pp, $7 (P). [ISBN: 0-486-24004-5] Republication of 1947 edition. Covers calculus of propositions, functions and classes. One-fourth of book devoted to analysis of conversational language. Appendix contains exercises for most sections and solutions. KS

Foundations, T(14-16: 1), S, L. Natural Logic. Neil Tennant. Edinburgh U Pr, 1978, 196 pp, $13. [ISBN: 0-85224-347-2] Introductory mathematical logic text for philosophy or mathematics majors which "attempts to reach major results with a minimum of fuss." Develops predicate calculus using natural deduction. Includes compactness, completeness, categoricity, incompleteness and undecidability for arithmetic, and a little recursion theory. Integrates discussion of classical and intuitionistic logic. 29 exercises. KS

Calculus, T(14: 1). Multivariable Calculus. James F. Hurley. Saunders College Pub, 1981, xiii + 576 pp, $21.95. [ISBN: 03-058604-6] This is the same text as Hurley's Intermediate Calculus (TR, June-July 1980) with the chapters on differential equations deleted. All comments still apply--it is an excellent text for a sophomore level multivariable calculus course. LLK

Real Analysis, S(16-17). Exercices et Problèmes d'intégration. Claude George. Gauthier-Villars, 1980, x + 432 pp, (P). [ISBN: 2-04-011246-4] A collection of problems-with-solutions for a standard course in integration. Topics include measurable sets, Fubini's theorem, functions of bounded variation, and trigonometric series. SES

Real Analysis, T(16-17: 1). L'intégrale. Paul Deheuvels. Pr U France, 1980, 228 pp, (P). [ISBN: 2-13-036648-1] A development of integration via integrals--the Riemann integral, Daniell integral, Lebesgue, Stieljes, Radon integrals, and some numerical methods. Includes exercises. SES

Differential Equations, P. Analytical and Numerical Approaches to Asymptotic Problems in Analysis. Ed: O. Axelsson, L.S. Frank, A. van der Sluis. Math. Stud., V. 47. North-Holland, 1981, xvi + 381 pp, $53.75 (P). [ISBN: 0-444-86131-9] Proceedings of a conference held June 9-13, 1980 at the University of Nijmegan, The Netherlands. It includes the texts of 17 invited addresses and 11 contributed papers which report on current pure and applied work in asymptotic analysis. AO

Functional Analysis, T, S(16-18), P, L. History of Functional Analysis. Jean Dieudonné. Math. Stud., V. 49. Elsevier North-Holland, 1981, vi + 312 pp, $29.50 (P). [ISBN: 0-444-86148-3] A comprehensive exposition of the evolution of modern functional analysis from late eighteenth century origins in differential equations and calculus of variations. Dieudonné paints the grand panorama, highlighting breakthroughs provided by key individuals and placing each in an appropriate mathematical setting. Spectral theory, duality and infinite dimensionality provide the major focal points; linear algebra and point set topology share Dieudonné's canvas, the roots of each being intimately linked to those of functional analysis. An excellent supplement to any standard course in functional analysis. LAS

Optimization, T*(15), L. Fundamentals of Management Science, Revised Edition. Efraim Turban, Jack R. Meredith. Business Pub, 1981, xviii + 651 pp, $22.50. [ISBN: 0-256-02393-X] The coverage of topics is very broad: decision theory, mathematical programming, queues, inventories, simulation. While the mathematical level is not high, the text is appealing for its many examples, exercises and applications. TAV

Optimization, P. Stochastic Programming. Ed: M.A.H. Dempster. Academic Pr, 1980, xiii + 573 pp, $111. [ISBN: 0-12-208250-8] Proceedings of an international conference at Oxford, July 1974. After a very readable introduction to the subject by the editor, 30 papers are presented about equally divided between theory and applications. The price is ridiculous! TAV

Analysis, P, L. Recurrence in Ergodic Theory and Combinatorial Number Theory. H. Furstenberg. Princeton U Pr, 1981, xi + 202 pp, $19.50. A beautiful exposition of the connection between aspects of dynamical system theory (recurrence, mixing, invariant measures) and the existence of arbitrarily long arithmetic progressions in certain sets of natural numbers. Could be used in a faculty seminar. SG

Analysis, P. Théorie Ergodique. L'Enseignement Math, 1981, 112 pp, Frs. 34 (P). A collection of seven papers covering such topics as entropy, flows to group actions on manifolds, etc. SG

Analysis, T*(16-17), L**.** Introduction to Classical Real Analysis. Karl R. Stromberg. Wadsworth, 1981, ix + 575 pp, $29.95. [ISBN:0-534-98012-0] A wonderful text, very thoughtfully constructed. A two-semester course from this text should be great preparation for graduate complex or abstract analysis. Large, interesting problem sets highlight each chapter. This could very well become a classic. TAV

Geometry, T(16-17: 1), S, P, L. Metric Planes and Metric Vector Spaces. Rolf Lingenberg. Wiley, 1979, xi + 209 pp, $25.50. [ISBN: 0-471-04901-8] This text develops and verifies the relations between a purely geometric treatment of Euclidean and non-Euclidean planes based on incidence structures and reflections. It also treats the theory of metric vector spaces and a theory of special types of S-groups. JNC

Geometry, S*(16-18), P*, L*. Geometry of Complex Numbers: Circle Geometry, Moebius Transformation, Non-Euclidean Geometry. Hans Schwerdtfeger. Dover, 1979, xiii + 200 pp, $4 (P). [ISBN: 0-486-63830-8] Uses portions of the theory of functions of a complex variable to develop understanding of the geometrical theory of analytic functions and the interrelatedness of geometry, analysis and algebra. Prerequisites: algebra of complex numbers, analytic geometry and linear algebra. A corrected republication of the 1962 edition. JNC

Geometry, T*(15-17: 1). An Outline of Projective Geometry. Lynn E. Garner. North Holland, 1981, vii + 220 pp, $29.50. [ISBN: 0-444-00423-8] A lucid, well-organized presentation of the fundamentals using the modern abstract point of view with analytic geometry illustrations. Requires some familiarity with analytic geometry, linear algebra and modern abstract algebra; the essential algebra topics are covered in an appendix. JNC

Combinatorics, T(16-18: 2, 3). Graphs, Networks, and Algorithms. M.N.S. Swamy, K. Thulasiraman. Wiley, 1981, xviii + 592 pp, $37.50. [ISBN: 0-471-03503-3] A thorough, book-length introduction to graphs and matroids followed by three substantial chapters on electrical network theory and two more on algorithmic graph theory. The latter treats both algorithmic analysis and algorithmic optimization. SS

Combinatorics, T(16-18: 1, 2), S, P. Applications of Graph Theory. Ed: Robin J. Wilson, Lowell W. Beineke. Acad Pr, 1979, xii + 423 pp, $49. [ISBN: 0-12-757840-4] Thirteen expository essays on applications of graph theory to communication and electrical networks, statistical mechanics, chemistry, social science, geography, architecture, linguistics, optimization, and computer science. SS

Combinatorics, P. The Interval Function of a Graph. H.M. Mulder. Math. Centre Tracts, No. 132. Math Centrum, 1980, iii + 191 pp, Dfl. 24,15 (P). [ISBN: 90-6196-208-0] A vertex x is between vertices u and v if x lies on a shortest path from u to v; the interval between u and v is the set of all vertices between u and v. The notion of interval is related to a variety of combinatorial and algebraic objects of interest that are explored in this tract. Among them are Hadamard graphs, median graphs, Hamming graphs, Helly hypergraphs, block designs and lattices. Careful exposition. SS

Number Theory, T(17-18), L. p-adic Numbers and Their Functions, Second Edition. Kurt Mahler. Cambridge U Pr, 1981, xi + 320 pp, $45. [ISBN: 0-521-23102-7] A vastly expanded second edition (First Edition, TR, June-July 1973). Extensive discussion of p-adic and g-adic functions: interpolation and approximation, derivatives, and integrals. Many good exercises. Unfortunately the price will place the book out of the reach of most mathematicians and libraries. SG

Number Theory, S(10-14), L. Have Some Sums to Solve: The Compleat Alphametics Book. Steven Kahan. Baywood Pub, 1978, 114 pp, $4.95 (P). [ISBN: 0-89503-007-1] 20 narrative alphametrics plus 20 ideal doubly-true alphametrics (whose verbalization is also true), with hints on how to get started in a middle section, and answers in the back. Scattered throughout is a Ripley's Believe-it-or-Not collection of "fabulous facts" about numbers. LAS

Linear Algebra, T(14: 1). Elementary Linear Algebra, Third Edition. Howard Anton. Wiley, 1981, xvii + 374 pp, $19.95. [ISBN: 0-471-05338-4] Additions to the Second Edition (TR, First Edition, March 1973; ER, March 1974; Second Edition, TR, May 1977) include supplementary exercises following Chapters 1, 2, 4, 5, and 6, and a new section on geometry of linear transformations. LLK

Algebra, T(18: 1), S, P. Hopf Algebras. Eiichi Abe. Trans: Hisae Kinoshita, Hiroko Tanaka. Tracts in Math., No. 74. Cambridge U Pr, 1980, xii + 284 pp, $39.50. [ISBN: 0-521-22240-0] Introduction to basic theory of Hopf algebras with emphasis on applications to affine algebraic groups and field theory. First chapter and appendix summarize background results on modules, algebras and categories. Some exercises. Revised and translated version of 1977 Japanese edition. KS

Algebra, P. Ordered Groups: Proceedings of the Boise State Conference. Ed: Jo E. Smith, G. Otis Kenny, Richard N. Ball. Lect. Notes in Pure and Appl. Math., V. 62. Dekker, 1980, x + 174 pp, $25.50 (P). [ISBN: 0-8247-6943-0] 14 papers from a conference held at Boise, Idaho, October 16-20, 1978. Several papers use ideas from mathematical logic and model theory in study of lattice-ordered groups. KS

Algebra, P. Abstract Witt Rings. Murray Marshall. Pure and Appl. Math., No. 57. Queen's U, 1980, vi + 257 pp, (P). Exposition of theory of abstract Witt rings, introduced in 1970's as abstract setting in which to study quadratic forms over a field. Includes applications of theory to quadratic forms over a semi-local ring. Uses equivalence between category of abstract Witt rings and category of quaternionic structures. KS

Algebra, P*, L. The Santa Cruz Conference on Finite Groups. Ed: Bruce Cooperstein, Geoffrey Mason. Proc. of Symp. in Pure Math., V. 37. AMS, 1980, xviii + 634 pp, $39.60. [ISBN: 0-8218-1440-0] Classification theory, general theory, properties of special groups, representation theory of groups of Lie type, character theory for finite groups, relations with combinatorics, computer applications, and connections with number theory: 90 papers from a 1979 summer research symposium showing that "far from being dead, group theory has only just come of age." LAS

Calculus, T(13: 1). A Short Course in Calculus with Applications to Management, Life and Social Sciences. Bodh R. Gulati. Dryden Pr, 1981, 536 pp, $17.95. [ISBN: 0-03-047466-3] This carefully written text covers the standard topics in first semester calculus, plus methods of integration and a chapter on functions of several variables. Most proofs are omitted, but concepts are well-motivated by examples and illustrations. Good selection of exercises to reinforce technique and illustrate applications. JRG

Calculus, T(13-14: 3). Calculus: Concepts and Calculations. A.W. Goodman, E.B. Saff. Macmillan, 1981, xix + 1028 pp, $28.95. [ISBN: 0-02-344740-0] Another 1000-plus page text for standard three-semester calculus course. Readable, conversational style. Includes proofs of most theorems. Some theoretical and calculator exercises. Usual applications to natural sciences. Sequences introduced early. Only separable differential equations. Answers to odd-numbered exercises. Portions adapted from Goodman's Analytic Geometry and the Calculus. KS

Geometry, S(15-17), L*. The Geometrical Foundation of Natural Structure: A Sourcebook of Design. Robert Williams. Dover, 1979, xv + 265 pp, $6 (P). [ISBN: 0-486-23729-X] A corrected republication of the 1972 volume Natural Structure (Eudaemon Pr., Moorpark, CA), this profusely illustrated typescript book provides a detailed resource of theoretical information (angles, volumes, networks) for tessellations, packings and arrangements. Williams' intent is to develop a language of form for artists, designers, scientists and engineers: a set of geometric entities with rules for their combination that can transform ideas and generate behavior. The elements of this form-language are taken primarily from nature--physics, botany, chemistry, metallurgy, psychology--and revealed in geometry. LAS

Geometry, S(9-10), L. Remarkable Curves. A.I. Markushevich. Trans: Yu. A. Zdorovov. MIR Pub, 1980, 47 pp, $2 (P). A brief non-analytic exposition for high school students of curves and their properties--from conic sections to leminiscates, spirals, and the catenary. An excellent supplement to school mathematics that could introduce students to special topics of beauty and utility. LAS

Geometry, T(13: 1), S. Analytical Geometry. A.V. Pogorelov. Trans: Leonid Levant. MIR Pub, 1980, 240 pp, $4.40 (P). A lucid, straightforward exposition of classical analytic geometry--from Cartesian coordinates through quadric surfaces and linear transformations. Concludes with solutions and hints to the numerous exercises. Published in the Soviet Union, this concise, compact (200 cm^3) volume--rare in the U.S. market--is a bargain at the price. LAS

Statistics, T(13: 1, 2). Descriptive and Inferential Statistics, An Introduction, Second Edition. Herman J. Loether, Donald G. McTavish. Allyn, 1980, xii + 659 pp, $18.95. [ISBN: 0-205-06905-3] Elementary statistics for sociology students, roughly 60% descriptive and 40% inferential. Changes have been made in all the chapters, and some computational problems have been added to the discussion questions of the 1974 first edition. RSK

Statistics, P*. COMPSTAT 1980: Proceedings in Computational Statistics. Ed: M.M. Barritt, D. Wishart. Physica-Verlag, 1980, 632 pp, (P). [ISBN: 3-7908-0229-8] Contains 85 papers on statistical computing presented at the 4th COMPSTAT Symposium held at Edinburgh in 1980. In addition to 4 invited lectures, papers are in the areas of sampling methods (3), data base management (9), education (3), analysis of variance and covariance (6), interactive computing (9), linear and nonlinear regression (6), multivariate analysis (6), optimization and simulation (6), cluster analysis (9), statistical software (18), and time series analysis (6). RSK

Statistics, P*. Abstract Inference. Ulf Grenander. Wiley, 1981, ix + 526 pp, $35. [ISBN: 0-471-08267-8] In the Wiley Series in Probability and Mathematical Statistics. Presumes a solid background in general probability theory, statistical theory, and functional analysis. Part I discusses prerequisite probability theory on abstract spaces. Part II presents an outline of the well-developed theory of inference in abstract sample spaces as it exists today. Part III presents the beginnings of a theory of inference when the parameter space is abstract. Good bibliography. RSK

Statistics, T(17-18: 1, 2), S, P, L. Mathematical Programming in Statistics. T.S. Arthanari, Yadolah Dodge. Wiley, 1981, xviii + 413 pp, $28.95. [ISBN: 0-471-08073-X] "An attempt to bring together most of the available results on applications of mathematical programming in statistics." Includes applications to cluster analysis. No exercises. Extensive bibliographies. FLW

Statistics, T*(13: 1), S, L. Elementary Statistics. Bernard W. Lindgren, Donald A. Berry. Macmillan, 1981, xi + 530 pp, $18.95. [ISBN: 0-02-370790-9] Traditional mathematical treatment of statistical topics; more informal discussion of probability. Many examples motivate and follow up new concepts. Particular attention to sampling problems and the abuse of statistical methods. JRG

Statistics, P. Analysis with Standard Contagious Distributions. J.B. Douglas. Intern Co-op Pub, 1980, xiv + 520 pp, $35. [ISBN: 0-89974-012-X] Volume 4 of the Statistical Distributions in Scientific Work Series. Study of some contagious distributions--less simple discrete distributions obtained, e.g., as mixtures of distributions or by taking a sum of a random number of random variables (stopped distributions). Emphsizes the Neyman Type A distribution, with less attention to the Poisson-binomial, Poisson-Pascal, Polya-Aeppli, Thomas, and log-zero-Poisson distributions. RSK

Statistics, T(14-17: 2), L. Statistics, Third Edition. William L. Hays. HR&W, 1981, xi + 723 pp, $22.95. [ISBN: 0-03-056706-8] Major revision of the author's 1973 text, Statistics for the Social Sciences, Second Edition (TR, October 1974). Much material has been eliminated, including the chapters on sets and functions and on Bayesian methods, while other topics have been expanded, such as the general linear model. Includes more and simpler problems. RSK

Statistics, T(16-17: 1), P. Digital Foundations of Time Series Analysis: Volume 1, The Box-Jenkins Approach. Enders A. Robinson, Manuel T. Silvia. Holden-Day, 1979, viii + 451 pp, $35. [ISBN: 0-8162-7270-0] Self-contained development and explanation of the basic principles of the Box-Jenkins method of time series analysis. Assumes background in matrix algebra (reviewed in an appendix) and mathematical statistics. Provides background for understanding the Box and Jenkins classic, Time Series Analysis: Forecasting and Control. (Revised Edition, TR, January 1977.) RSK

Statistics, P*. Statistical Inference for Stochastic Processes. Ishwar V. Basawa, B.L.S. Prakasa Rao. Prob. and Math. Stat. Acad Pr, 1980, xiv + 435 pp, $71.50. [ISBN: 0-12-080250-3] Divided into three roughly equal parts. Part I discusses standard models, including discrete and continuous Markov chains and simple point processes. Part II provides the theory of inference for general processes in discrete and continuous time, including diffusion processes. Part III surveys recent results on Bayesian, non-parametric, and sequential procedures. Good set of references. RSK

Statistics, T(18: 1), P. The Theory of Linear Models and Multivariate Analysis. Steven F. Arnold. Wiley, 1981, xv + 475 pp, $34.95. [ISBN: 0-471-05065-2] In the Wiley Series in Probability and Mathematical Statistics. Detailed theoretical treatment of models which assume an underlying normal distribution, including the univariate linear model, the generalized linear model, the repeated measures model, random effects and mixed models, the correlation model, the multivariate one- and two-sample models, the multivariate linear model, and the discrimination model. Presumes a solid background in mathematical statistics and matrix algebra. RSK

Computer Science, P. Correctness Preserving Program Refinements: Proof Theory and Applications. R.J.R. Back. Math. Centre Tracts, No. 131. Math Centrum, 1980, 118 pp, Dfl. 15,75 (P). [ISBN: 90-6196-207-2] Monograph describing use of infinitary logic to prove correctness of refinement steps in program construction. Introduces a binary refinement relation such that relation holds if and only if a corresponding formula of infinitary logic is provable. KS

Computer Science, P. Mathematical Methods in Computer Graphics and Design. Ed: K.W. Brodlie. Acad Pr, 1980, xi + 147 pp, $26. [ISBN: 0-12-134880-6] Proceedings of a conference held at the University of Leicester on September 28, 1978. Contains six papers on curve and function drawing, contouring, and computer-aided design. AO

Computer Science, T(14-15: 1). File Techniques for Data Base Organization in COBOL. LeRoy F. Johnson, Rodney H. Cooper. P-H, 1981, xvi + 368 pp, $19.95. [ISBN: 0-13-314039-3] Textbook for a second course on COBOL programming with an emphasis on basic file processing techniques. The final chapter includes a discussion of the CODASYL COBOL data base facility. AO

Systems Theory, P. Identification of Control Plants. Zdzisaw Bubnicki. Stud. in Automation and Control, V. 3. Elsevier Scientific Pub, 1980, xi + 312 pp, $68.25. [ISBN: 0-444-99767-9] This monograph presents a unified treatment of the field of system identification (i.e., the determination of mathematical models ot processes on the basis of empirical data) while focusing on the identification of control plants (i.e., processes submitted to control). AO

Applications (Biology), S(13-15), L. The Curves of Life. Theodore Andrea Cook. Dover, 1979, xxx + 479 pp, $5.95 (P). [ISBN: 0-486-23701-X] Unabridged republication of a classic compendium on the golden ratio, first published in London in 1914. Eleven plates and 415 illustrations depict spiral patterns in shells, plants, flowers, horns, anatomy, architecture and art; extensive commentary exudes neo-Newtonian awe in the ability of mathematical laws of growth to approximate living reality. LAS

Applications (Biology), P. Lecture Notes in Biomathematics-39: Vito Volterra Symposium on Mathematical Models in Biology. Ed: Claudio Barigozzi. Springer-Verlag, 1980, vi + 417 pp, $28 (P). [ISBN: 0-387-10279-5] Proceedings of a December 1979 conference held at the Accademia Nazionale dei Lincei in Rome, in two parts: models of natural selection, and problems of population biology. LAS

Applications (Physics), P. Mathematical Modelling of Turbulent Diffusion in the Environment. Ed: C.J. Harris. Acad Pr, 1979, xvi + 500 pp, $48. [ISBN: 0-12-328350-7] Eight papers on the dispersion of pollutants in air and six on water pollution from a September 1978 conference at Liverpool University sponsored by the Institute of Mathematics and Its Applications. LAS

Applications (Political Theory), T, S*(13-15), L*. Topics in the Theory of Voting. Philip D. Straffin, Jr. Birkhäuser Boston, 1980, ix + 69 pp, $5 (P). [ISBN: 3-7643-3017-1] A monograph from UMAP (Undergrduate Mathematics and its Applications Project) surveying three aspects of the theory of voting: measures of voting power, voting methods when there are more than two alternatives, and voting schemes based on intensities of preference. Requires only minimal mathematics background; ideal resource for freshman seminar or introductory course in mathematical modelling. LAS

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

Readers are invited to contribute to the general interest of this department by sending news items to Mathematical Association of America, 1529 Eighteenth St., N.W. Washington, D.C. 20036

PERSONAL ITEMS

George H. Andrews has been named Chairman of the Department of Mathematics at Oberlin College for the period February 1, 1981 to June 30, 1984.

Arthur P. Boblett has retired from the Naval Weapons Center, China Lake, CA, after thirty-one years of Federal services.

Instructor *LeRoy P. Hammerstrom* has been promoted to Assistant Professor at Eastern Nazarene College, Quincy, MA.

Anthony G. O'Farrell of Maynooth College, County Kildare, has been elected a member of the Royal Irish Academy.

Paul Olum, for many years a professor of Mathematics at Cornell University and more recently Vice-President for Academic Affairs at the University of Oregon, has been named President of that institution.

Warren Page, New York City Technical College, was presented with The Distinguished Achievement Award for Mathematics Education by the New York State Mathematics Association of Two-Year Colleges at its 16th annual meeting, Niagara Falls, N.Y.

Kenneth B. Reid, Louisiana State University, has been chosen for the 1981 class of the W.K. Kellogg Foundation's National Fellowship Program. Forty outstanding young American professionals were selected for the purpose of increasing their leadership abilities in order to deal more effectively with society's problems.

B.M. Stewart has retired with the title of Professor Emeritus after forty years of teaching at Michigan State University

Hugo M. Beck, formerly a mathematician with the U.S. Navy, died March 4, 1981. He was a member of the Association for eleven years.

Henry J. Miles, Professor Emeritus at the University of Illinois, died November 13, 1980. He was a member of the Association for forty-nine years.

Halvor T. Darracott of Falls Church, VA, died April 1, 1981. He was a member of MAA since 1942.

The deaths of these members of MAA have recently been reported: Dr. *H.W. Raudenbush* of Floral Park, N.Y., in April 1978; *Walter O. Shriner* of Terre Haute, Indiana, and *Willis B. Caton*, Professor Emeritus at Illinois Institute of Technology.

COMING SOON IN THE MONTHLY

The following articles will appear in the AMERICAN MATHEMATICAL MONTHLY for November 1981:

M. Katchalski, M.S. Klamkin, and *A. Liu,* An Experience in Problem Solving
W.B. Arthur, Why a Population Converges to Stability
Karen D. Rappaport, S. Kovalevsky: A Mathematical Lesson
William Abikoff, The Uniformization Theorem
Patricia C. Kenschaft, Black Women in Mathematics in the United States
L.F. Klosinski, G.L. Alexanderson, and *A.P. Hillman,* The William Lowell Putnam Mathematical Competition

The following articles are among those which have been accepted by the editors for later issues of the MONTHLY. The order of listing does not indicate the order in which they will appear.

Geoffrey C. Berresford, Cauchy's Theorem
R.P. Boas, Can We Make Mathematics Intelligible?
Edwin Buchman, The Impossibility of Tiling a Convex Region with Unequal Equilateral Triangles
Carmen Chicone and *Tian Jinghuang,* On General Properties of Quadratic Systems
A.J.W. Duijvestijn, P.J. Federico and *P. Leeuw,* Compound Perfect Squares
Solomon W. Golomb, Irrational Sums and Twin Primes
Heinrich Guggenheimer, The Hilbert Model of Hyperbolic Geometry
Richard K. Guy, Research Problems Become Unsolved Problems
Li Hong-Xiáng, Elementary Quadratures of Differential Equations
Donald E. Knuth, A Permanent Inequality
Katherine E. McLain and *Hugh M. Edgar,* A Note on Golomb's "Cyclotomic Polynomials and Factorization Theorems"
Samuell Merrill, III, Approximations to the Banzhaf Index of Voting Power
F.H. Norwood, Long Proofs
Robert C. Reilly, Mean Curvature, the LaPlacian and Soap Bubbles
Tony Rothman, The Fictionalization of Evariste Galois
Wolfgang Walter, A New Approach to Euler's Trigonometric Expansions
John J. Wavrik, Computers and the Multiplicity of Polynomial Roots

OFFICIAL ANNOUNCEMENTS AND COMMUNICATIONS

ANNOUNCEMENT OF ALLENDOERFER, FORD AND POLYA AWARDS

At its meeting on January 28, 1977, in St. Louis, Missouri, the Board of Governors authorized a number of awards to authors of expository articles published in the MONTHLY, to be named after Lester R. Ford, Sr., MATHEMATICS MAGAZINE, to be named after Carl B. Allendoerfer, and the TWO-YEAR COLLEGE JOURNAL, to be named after George Polya. A maximum of two Carl B. Allendoerfer Awards, five Lester R. Ford Awards, and two George Polya Awards will be made annually; each award is in the amount of $100. The articles are to be selected by committees appointed by the President of the Association for this purpose and the Chairman of the Committee on Publication is to be an *ex-officio* member of each of these committees.

The recipients of the Carl B. Allendoerfer Awards for 1980 were selected by a committee consisting of Roy Dubisch, Chairman; Edwin F. Beckenbach, *ex-officio;* and Thomas W. Tucker. The recipients for Allendoerfer Awards for articles published in 1980 were the following:

Stephen H. Maurer, "The King Chicken Theorems," MATHEMATICS MAGAZINE 53 (1980), 67-80
Donald E. Sanderson, "Advanced Plane Topology from an Elementary Standpoint," MATHEMATICS MAGAZINE 53, 81-89

The recipients of the Lester R. Ford Awards for 1980 were selected by a committee consisting of Branko Grunbaum, Chairman; Edwin F. Beckenbach, *ex-officio*; and Peter L. Duren. The recipients for Ford Awards for articles published in 1980 were the following:

Lawrence Zalcman, "Offbeat Integral Geometry," MONTHLY 87 (1980), 161-175
R. Creighton Buck, "Sherlock Holmes in Babylon," MONTHLY 87 (1980), 335-345
B.H. Pourciau, "Modern Multiplier Rules," MONTHLY 87 (1980), 433-452
E.R. Swart, "The Philosophical Implications of the Four-Color Problem," MONTHLY 87 (1980), 697-707
Alan H. Schoenfeld, "Teaching Problem-Solving Skills," MONTHLY 87 (1980), 794-805

The recipients of the George Polya Awards for 1980 were selected by a committee consisting of Kay W. Dundas, Chairman; Edwin F. Beckenbach, *ex-officio;* and Warren Page. The recipients for Polya Awards for articles published in 1980 were the following:

E.D. McCune, R.G. Dean, and W.D. Clark, "Calculators to Motivate Infinite Composition of Functions," Volume II, Number 3, Pages 189-195, TWO-YEAR COLLEGE MATHEMATICS JOURNAL
G.D. Chakerian, "Circles and Spheres," Volume II, Number 1, Pages 26-41, TWO-YEAR COLLEGE MATHEMATICS JOURNAL

1981 CONTRIBUTING MEMBERS AND SPECIAL GIFTS

The Association is deeply indebted to the generosity of the 91 members listed below who have elected to be Contributing Members, Sponsors or Patrons for 1981 by making contributions beyond the normal dues and to the 1 member who elected to make a special contribution.

Special Gift: Harry M. Gehman

Patrons: William G. Chinn, Addison M. Fischer, Bill Hassinger, Jr., Herbert J. Ryser, Saturnino L. Salas, Earl W. Swokowski

Sponsors: Barbara J. Beechler, Malcolm K. Brachman, George H. Bridgeman, Richard A. Cleveland, Harvey S. Fox, Andrew M. Gleason, Melvin Hochster, Peter Henrici, James E. Kiefer, Donald L. Kreider, Henry A. Krieger, James A. Long, Gene M. Ortner, H.O. Pollak, Robert H. Sorgenfrey

Contributing Members:

Henry M. Alder
Tom M. Apostol
Bernice L. Auslander
George A. Baker
Thomas W. Bartenstein
James R. Baugh
William M. Boyce
Mats H. Brobert
Ezra A. Brown
Alfred H. Clifford
A.J. Coleman
Henry D. Colson
Lewis C. Corey
James P. Crawford
Wilbur E. Deskins
William J. Dodge
James A. Donaldson
Kalus Eldridge
Earl O. Embree
Ian M. Ferris
Richard L. Gantos
Milton A. Glass
Richard Goldberg
William A. Golomski
W.R. Harris
William L. Hart
William E. Hartnett
Anna S. Henriques
Paul S. Herwitz
Herbert L. Holden
John M. Horvath
Roy A. Johnson
Kenneth E. Kloss
Anne B. Koehler
Ralph M. Krause
Joseph P. Lasalle
Robert N. Leggett
Bernard W. Levinger
John A. Lewis
Viktors Linis
W.R. Mann
Maynard J. Mansfield
John R. Mayor
Robert K. McConnell, Jr.
Robert F. McNaughton, Jr.
E.P. Miles
Peter L. Montgomery
Richard A. Moore
Somashekha A. Naimpally
Dale A. Nelson
Cecil J. Nesbitt
Barbara L. Osofksy
David E. Penny
John M. Perry
Joel Pitcairn
Kenneth R. Rebman
Donald C. Rose
David P. Roselle
James R. Schap
Norman E. Sexauer
Juriaan Simonisio
David B. Singmaster
Joel B. Starkey
Andrew J. Sterrett
Carl R. Tellefsen
Samuel S. Wagstaff, Jr.
J. E. Wilkins
Elizabeth B. Weinstock
Lawrence A. Salcman
Robert E. Zink

THE 1981 WILLIAM LOWELL PUTNAM MATHEMATICAL COMPETITION

The 42nd Annual William Lowell Putnam Mathematical Competition will be held at participating institutions on Saturday, December 5, 1981. This competition is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship and is administered by The Mathematical Association of America. All colleges and universities in Canada and the United States may register eligible undergraduates. Registration forms will be mailed to institutions that participated in the 41st competition by September 22, 1981. Other institutions that wish to enter undergraduates should request registration forms from Professor L.F. Klosinski, Director; The William Lowell Putnam Mathematical Competition; University of Santa Clara; Santa Clara, CA 95053. Completed registrations must be received by the Director no later than October 16, 1981.

Further details are given in the Announcement Brochure that is mailed with the registration material. Reports of previous competitions, including examination questions and outlines of solutions, are in past issues of this MONTHLY: the most recent of these reports were in the issues of October 1980, November 1979, March 1979, January 1978 and November 1976.

SOUTHERN CALIFORNIA SECTION MEETING

The annual Spring Meeting of the Southern California Section of the MAA was held March 7, 1981 at California State University, Long Beach. Approximately 115 people registered.

At the business meeting election results were announced and John Todd, Governor of the section-reported on the January Governors' Meeting. The following invited addresses were well received: *George Sell,* University of Minnesota and USC, "Strange Attractors in Dynamical Systems:" *David Wales,* Caltech, "Simple Groups; Their Classification and Some Consequences;" *Robert D. Edwards,* UCLA, "Linkages and Manifolds;" *W. Gilbert Strang,* MIT and Caltech, "Optimal Design in Engineering and Biology."

At the luncheon, *F. Burton Jones,* U. of Cal., Riverside, gave a talk entitled, "Texas Mathematical Humor." Following the luncheon a panel discussed "Precollege Mathematics Skills: Can a Cooperative High School-University Effort Reverse the Decline?" The members of the panel were: *Phil Curtis,* UCLA, Moderator; *Robert Blackburn,* Washington High School; *Merton Burkhard,* Santa Monica High School and UCLA; *Gerald Marley,* Cal State University, Fullerton; *Herbert Niebergall,* Fremont High School.

KENTUCKY SECTION MEETING

The sixty-fourth annual meeting of the Kentucky section of the MAA was held at Jefferson Community College, Louisville, KY., on April 3-4, 1981.

Invited speakers were Prof. *R. P. Boas,* editor of the AMERICAN MATHEMATICAL MONTHLY, and Prof. *Jerry King,* Lehigh University. On Friday evening, Prof. King spoke on "Intuition and Elegance" and Prof. Boas spoke on "Some Heretical Thoughts on Teaching Mathematics." On Saturday, Prof. King's topic was "Approximations of Continuous Functions" while Prof. Boas' topic was "Serious Mathematics from Unpromising Material."

Contributed papers given on Saturday were: "On Formulating the Hyperbolic Functions," Prof. *James Barksdale,* Western Kentucky University; "Micro Calculus: Calculus on a Micro Computer," Prof. *George Barnes,* University of Louisville; "An Algorithm for Graphing Implicit Functions, Using Newton's Method," Prof. *Robert Lindahl,* Morehead State University; "Some Algebraic Properties of the Convolution Integral," Prof. *Joseph Stokes,* Western Kentucky University; and, "Video Enhancement of Micro Computers in Mathematics Education," Prof. *Roger Geeslin,* University of Louisville.

Prof. *Martha Watson,* Western Kentucky University, was elected section governor for 1981-84 and Prof. *Peter Moore,* Northern Kentucky University, was elected Chairman-Elect for 1981-83. Prof. *Richard Davitt,* University of Louisville, assumed the chairmanship for 1981-83.

The sixty-fifth meeting will be held at the University of Kentucky, Lexington, April 2-3, 1982.

SOUTHWESTERN SECTION SPRING MEETING

The 41st annual meeting of The Southwestern Section MAA was held at New Mexico State University, Las Cruces, New Mexico, on April 3-4, 1981. Approximately 60 members and guests registered their attendance.

MAA President *R.W. Anderson,* Louisiana State University, gave the banquet address entitled, "Washington Representation of Mathematics." He also gave an invited address at the Saturday morning session entitled, "Some Elementary Ideas in Infinite Dimensional Topology."

Professor *Ruth Rebekka Struik,* University of Colorado-Boulder, gave two invited addresses entitled "Women Mathematicians, Some Famous and Some Not So Famous," and "A Group Construction Associated with LaGrange's Theorem."

Professor *Jean J. Pederson,* University of Santa Clara, also gave two invited lectures. The titles were "There Is More to Geometric Figures than Meets the Eye," and "Teaching Mathematics to Adults."

The sessions for contributed papers were chaired by Professors *Elbert Walker, Judy Moore, Edward Gaughan, Don Johnson,* and *Robert Wisner.* The following papers were contributed: "Functional Iteration of the Unit Interval, Recent Developments," *William A. Beyer,* Los Alamos National Laboratory; "Supply, Demand and Rearrangements," *Richard Bagby,* New Mexico State University; "Force is a Line-Bound Vector," *A. Swimmer,* Arizona State University; "A Lucas Type Condition for the Primality of Fermat Numbers, Part II," *Justin G. MacCarthy,* Demind New Mexico; "Finitely Generated Subspaces of One-Dimensional Real Vector Spaces," *Fred Richman,* New Mexico State University; "An Historical Over-View of a Developmental Math Course," *Carmen A. Leal,* El Paso Community College; "Department of Developmental Studies--An Approach to Remediation," *Peter B. Bonner,* Albuquerque Technical-Vocational Institute; "Placement and Remediation at New Mexico State University," *Don Johnson,* New Mexico State University

On Saturday morning there was a panel discussion on remedial instruction. The panelists were: *Laura Cameron,* University of New Mexico; *Ron Loser,* Adams State College; *Sam Self,* El Paso Community College. *Don Johnson,* Chairman of the Southwestern Section acted as meeting coordinator.

APRIL MEETING OF THE SEAWAY SECTION

The Seaway Section of MAA held its spring meeting at Syracuse University on April 10 and 11, 1981. There was a registered attendance of 87 individuals. Section Chairperson *Howard Bell* presided.

On Friday, the Executive Committee and Committee Chairpersons met to consider future activities of the Section. This was followed by a banquet and later in the evening a delightful talk entitled, "The Physics of Toys" by Professor *Henry Levinstein* of Syracuse University.

The Saturday sessions were devoted to the presentation of several papers by members of the Section and the Gehman Lecture of Professor *J.H.B. Kemperman* of the University of Rochester, entitled, "Systems of Mating and Equilibrium in the Presence of Imprinting,". Copies of Professor Kemperman's paper may be obtained from *D.W. Trasher*, the Section's Secretary-Treasurer.

Papers presented: "Foliations on Compact Riemannian Manifolds and the Ricci Curvature of the Leaves," *Richard Escobales*, Canisius College; "Sequences of Sets: Various Convergence Concepts in Topology," *Louis F. McAuley*, SUNY-Binghamton; "Perron-Frobenius Theory for Certain Affine Operators: An Application," *Kenneth Lane*, Hamilton College; "Luck as a Statistical Phenomenon," *Dennis S. Martin*, SUC-Brockport;"p-ADIC Solutions of $xyz = x + y + z = 1$: Propaganda for Quadratic Reciprocity," *Charles Small*, Queen's University; "Solving Linear Equations by Row and Column Reduction," *Norman Rice*, Queen's University; "Times Series in M Dimensions: An Introduction and Exposition," *Leo A. Aroian*, Union College.

At the business session *Irvin Jungreis* of Cornell University was recognized as the winner of the Section's annual award for the highest attainment on the Putnam Examination of students within the Section.

The Section elected *Robert Knapp*, Herkimer County Community College, to a two-year term as Second Vice-Chairman and *Donald Trasher*, SUC-Geneseo, to a second term as Secretary-Treasurer.

APRIL MEETING OF THE OHIO SECTION

The Ohio Section of MAA held its sixty-fifth annual Spring meeting at Miami University, Oxford, Ohio, April 10-11, 1981. One hundred and thirty people were present. Section Chairman *C. A. Long* presided; *D. J. Horwath* was the program chairman. The invited address, "The Harmonic Series and the Elephants," was presented by *R. P. Boas*, Northwestern University.

The following contributed papers were also presented: "What if...? An Exploration with Lines," *K.B. Cummins*, Kent State University; "Concurrent Teaching-Research-Consulting: A Unified Approach," *G.T. Frey*, Ohio State University-Marion; "The Mathematics of Gerrymandering," *C. R. Hampton*, College of Wooster; "Use of a Hand-Held Calculator for Solving Problems from the American Mathematical Monthly," *D. Moore*, Somerville, Ohio; "Numerical Approximations of Analytic Functions on an Ellipse," *T. E. Price, Jr.*, University of Akron; "Cubes of Sums," *L. D. Rodabaugh*, University of Akron; "The 1981 MAA Test-Problems and Solutions, *L. J. Schneider*, John Carroll University; "The Odds in a Pyramid Game--a Computer Simulation," *A. C. Stickney* and *E. L. Wilson*, Wittenberg University; "An Elementary Proof That Four Vectors in E^3 Are Linearly Dependent," *G. L. Szoke*, University of Akron.

The meeting agenda also included meetings of the Executive Committee and of ad hoc committees. Additional program highlights included: 'Biographies of Famous Mathematicians' Lecture, "Mathematicians of India," by *H. Shankar*, Ohio University; retiring chairman's address, "Singular Value Decomposition of Matrices with Applications," by *C.A. Long*, Bowling Green State Univ.; and the following special sessions: Student Papers, supervised by *M. Cox* (Miami Univ.); Swap Session: "Mathematics and the Adult Student, moderated by *J. A. Engle* (Ohio State Univ.); "Computer Graphics," moderated by *F. Shuurmann* (Miami Univ.); and meeting of The Association of Women in Mathematics, chaired by *J. A. Engle* (Ohio State Univ.)

The officership and committee chairmanships for the academic year 1981-82 include: Executive Committee; *J.D. Faires* (Youngstown State Univ.), Section Chairman; *D. J. Horwath* (John Carroll Univ.), Section Chairman-elect; *C. A. Long* (Bowling Green State Univ.) Section Past-chairman; *G. Mavrigian* (Youngstown State Univ.), Secretary-Treasurer; *A. Sterrett, Jr.* (Denison Univ.), Secretary-Treasurer-elect; *S. W. Hahn* (Wittenberg Univ.), Sectional Governor; and *A. G. Poorman* (Ashland College), Program Committee Chairman. Program Committee: *A. G. Poorman*, chairman; *J. P. Leitzel* (Ohio State Univ.); and *F.P. Merkes* (Univ. of Cincinnati). Nominations Committee: *D. O. Koehler* (Miami Univ.), Chairman; *W.H. Beyer* (Univ. of Akron); and *M.D. Wetzel* (Denison Univ.). Ad Hoc Committee Chairmen: Committee on Curriculum, *H. L. Putt* (Univ. of Akron); Committee on Teacher Training and Certification, *W. A. Kirby* (Bowling Green State Univ.); Committee on Section Activity, *P. H. Schmidt* (Univ. of Akron); Representative to the Two-year College Mathematics Journal, *C. P. Yang* (Miami Univ.-Middletown); High School Mathematics Competition Supervisor, *L. J. Schneider* (John Carroll Univ.); Public Informations Officer and Newsletter Editor, *R. A. Little* (Baldwin-Wallace College).

NEBRASKA SECTION

The Nebraska Section met on Friday and Saturday, April 10-11, 1981, in Vermillion, South Dakota, in the Churchill-Haines Laboratories on the University of South Dakota Campus. Invited lectures were given by Professors *Paul Campbell*, of Beloit College, and *Lynn Steen*, of St. Olaf College. In addition, Professor Alexander Mehaffey, Chairman of the Section, arranged for a visit to the Shrine of Music Museum and for interviews with Dr. *Andre P. Larson*, Director and Founder of the Museum. Attendance of forty or more was recorded for both sessions of the meeting. The program was as follows: "Equivalence of the Cantor-LeBesgue and Denjoy-Lusin Properties for Multivariable Trigonometric Series," *Vivian C. Mosca*, Univ. of South Dakota; "Applications of the Echelon Technique to Various Algebraic Algorithms," *L. M. Larsen*, Kearney State College; "Group Codes for the M-Receiver Gaussian Broadcast Channel," *John Karlof*, Univ. of Nebraska-Omaha; "Using an Auditing Language as an Information Retrival Tool," *John Powell*, University of South Dakota; "Rigor and Nature of Mathematics," *Leroy Meyer*, Univ. of South Dakota; "Finding the Smallest Primitive Root of a Prime," *James Hollerman*, Univ. of South Dakota; "Questions of Dependence for Meromorphic Mappings into Algebraic Varieties," *S. J. Drouilhet*, Yankton College; "Mathematics Tomorrow," *Lynn Steen*, First Vice-President of the MAA, St. Olaf College; lecture by *Paul Campbell* of Beloit College, "From the Prairie to Gottingen: The Life of Anna Pell Wheeler;" "Partial Differential Equation Problems as Encountered in Reservoir Simulation,"

Alan V. Lair, Univ. of South Dakota; "Components of Information for Categorical Data," *Gene B. Iverson*, Univ. of South Dakota; "Tooth Tables," *Gary Meisters*, Univ. of Nebraska-Lincoln; "The USD High School Mathematics and Computer Science Contest," *Andrew Karantinos*, Univ. of South Dakota; Report on the 1981 Annual High School Mathematics Contest," *Stanley B. Luke*, Nebraska Wesleyan Univ.; Business Meeting; "Integrating Computers into Undergraduate Mathematics," *Lynn Steen*, First Vice-President of the MAA, St. Olaf College.

Professor *Paul Campbell* provided a display of MAA Publications. Professor *Lynn Steen* reported on current interests and activities of the MAA.

Officers for 1981-82 were elected as follows: Chairman, *Randall K. Heckman*, Kearney State College; Chairman-elect, *John K. Karlof*, University of Nebraska at Omaha; Chairman of High School Mathematics Examination Committee, *Richard Barlow*, Kearney State College; Newspaper Editors, *Thomas S. Shores* and *Henry M. Cox*, University of Nebraska; Governor (1981-1983), *Gerald W. Johnson*, University of Nebraska; Secretary-Treasurer, *Henry M. Cox*.

APRIL 1981 MEETING OF THE SOUTHEASTERN SECTION

The sixtieth annual meeting of the Southeastern Section was held on April 10-11, 1981, at the University of Alabama in Birmingham. A total of 174 persons attended the meeting, including 18 students and 156 members. The local arrangements were handled by Professor *W. Fred Martens*.

Three invited addresses were given: Professor *R.D. Anderson* (President of the Association) of Louisiana State University on "Some Algorithmically Defined Functions;" Professor *B.F. Caviness* of Rensselaer Polytechnic Institute and General Electric Company on "A Survey of Symbolic Mathematical Computation;" and Professor *Daniel D. Warner* (Section Lecturer) of Clemson University on "Some Old and New Results on the Geometry of Polynomial Zeros."

There were nine sessions for contributed papers. The presiders were *Billy Grant* (Chairman of the Section), *John Neff* (Governor of the Section), and *W. Fred Martens* for the general sessions; and *John Baxley*, *Sara Ripy*, *Richard G. Vinson*, *Trevor Evans*, *Sandria Kerr*, *Thomas Blackburn*, *Fred Howard*, *John Kenelly*, and *Ray Wylie* for the special sessions.

Officers elected for 1981-82 are: Chairman, *Harvey Carruth*, University of Tennessee at Knoxville; Chairman-elect, *Mary E. Neff*, Emory University; Vice-Chairman, *Catherine C. Aust*, Clayton Junior College; Section Lecturer, *Ray Wylie*, Furman University--all for one year; and *Ivey C. Gentry*, Secretary-Treasurer for a term of three years.

At the business meeting, it was announced that the winner of the Section prize for the best performance on the Putnam examination was *Edward J. Rak* of the University of North Carolina at Chapel Hill.

The Section voted to hold its 1983 meeting at the Citadel, Charleston, South Carolina.

The following papers were presented: "The Sequence x_n Defined by $x_{n+1}= b^{(x_n)}$," *Benjamin G. Klein*, Davidson College; "Laplace Transforms That Don't Exist, or Do They?," *Brian M. O'Connor*, Tennessee Tech University; "Burnside's Formula for Factorial n," *Francis J. Murray*, Duke University; "A Humanities Approach to Mathematics," *B. F. Bryant*, Vanderbilt University; "A Statistical Study Aimed at Improving Placement in Freshman Mathematics," *Sandria N. Kerr*, Winston-Salem State University; "Rewriting Pre-calculus Mathematics," *Ray Wilson*, Central Piedmont Community College; "Quasi-partially Isometric Matrices," *Teh-Huey Chuang*, Tuskegee Institute; "Scalene Partitions of Positive Integers," *J. R. Ridenhour*, Austin Peay State University; "A Little League Schedule Using 1-factorization," *Tina H. Straley*, Kennesaw College; "The Mathematics Clinic at the University of Alabama," *James J. Buckley*, The University of Alabama; "Flow System Integrals," *Arthur C. Segal*, The University of Alabama; "How Many Steiner Triple Systems Are There?" *Curt Lindner*, Auburn University; "Numerical Analysis as a Focal Point in the First Two Years of College" *David Tudor*, The College of Charleston; "Mathematics in the Chinese Classroom--an Eyewitness Report," *James G. Ware*, University of Tennessee at Chattanooga; "Teaching Functions and Graphs--Some Interesting Questions," *James E. Bright*, Clayton Junior College; "Teaching Problem Solving--a Report on a LOCI Project," *David R. Stone*, Georgia Southern College; "An Alternate System of Instruction," *Thomas L. Alexander*, University of Alabama; "What on Earth is 1/0?" *Robert Tilidetzke*, Bennett College; "Computer Simulation of a Ski Resort," *Lyndell M. Kerley*, East Tennessee State University; "The Mathematical Paintings of Crockett Johnson," *J.B. Stroud*, Davidson College; "An Application of Dynamic Programming to a Problem in Forestry," *John I. Moore* and *James W. Urscheler*; "Using Tables to Solve Word Problems in Intermediate Algebra," *Catherine C. Aust*, Clayton Junior College; "On Cutting Twisted Prismatic Rings," *L. R. King*, Davidson College; "Non-existence of a Pair of Orthogonal Latin Squares," *Andrew Sobczyk*, Clemson University; "Beyond Pascal's Pyramid," *Arthur G. Sparks*, Georgia Southern College; "Algorithm for Evaluation of Determinants," *Larry R. Murdock* and *E. E. Moyers*, Integraph Corporation; "Table Tennis andPascal's Triangle," *John Neff*, Georgia Institute of Technology; "Some Remarks on Quaternionic Field Theory" *Frank D. (Tony) Smith, Jr.*, Cartersville, Georgia; "Data Considerations to Improve Statistical Forecasts," *Gaylord May*, Wake Forest University; "Limacon: a Case for a Fourth Case," *Linda H. Boyd* and *Charles R. Stone*, DeKalb Community College; "Color Graphics Reveal the Beauties of Mathematics," *E. P. Miles*, Florida State University; "Sequences Defined Recursively by $a_n = \sqrt{a_{n-1}} + \sqrt{a_{n-2}}$ Which Oscillate About Their Limit," *Elmer K. Hayashi*, Wake Forest University; "Convergence of $\Sigma s_n p^{a_n}$" *E.P. Turner*, Georgia State University; "A Mathematical Model for the Effect of Morphine," *John V. Baxley*, Wake Forest University; "Continuous Computation by Digital Algorithms," *Peter M. Winkler*, Emory University.

APRIL MEETING OF THE KANSAS SECTION

The sixty-sixth annual meeting of the Kansas Section of the MAA was held on April 10-11, 1981, at Benedictine College, Atchison, Kansas. Section Chairman *S. Thomas Parker* presided. Approximately 90 people attended.

Invited addresses were "Computers and Mathematical Proof" and "Proof of the Four Color Theorem," by *Kenneth I. Appel*, University of Illinois, and "Calculator Calculus at Ohio State," by *John Riedl*, Ohio State University.

The following contributed papers were presented: "The Math Anxiety Workshop--What We Learned," *Mary Ellen Foley* and *Donna Gorton*, Wichita State University; "A Fast GCD Algorithm," *Elwyn H. Davis*, Pittsburg State University; "Gamma Geometries Which Are Locally Polar Spaces," *Peter Johnson*, Kansas State University; "Tiltup Panels--Locate the Pulleys," *Gary McGrath*, Pittsburg State; "The Shape of the Strongest Fixed-Fixed Column," *Philip G. Kirmser*, Kansas State University; "Some Results and Open Questions Concerning Niven Numbers," *Robert E. Kennedy*, Central Missouri State University; "Comments on the Early History of Curvature," *Saul Stahl*, University of Kansas; "The Pigeon-Hole Principle," *Prem N. Bajaj*, Wichita State University; "Comparison of the Finite Element Method and the Least Squares Method," *Richard Lasseter*, Pittsburg State University.

THE REGULAR MEETING OF THE IOWA SECTION

The 68th regular meeting of the Iowa Section, MAA, was held on the campus of Coe College, Cedar Rapids, Iowa, on April 24-25, 1981. The meeting was held jointly with the Iowa sections of the Society for Industrial and Applied Mathematics and the American Statistical Association. While the Iowa Academy of Science was meeting at the same location, there was no formal connection with the MAA meeting. Professor *Donald Bailey* presided over the Friday session; Chairperson *Arnold Adelberg* presided over the Saturday norming session, and Chairperson-elect *A. M. Fink* and Professor *James Murdock* presided at the Saturday afternoon sessions. There were 24 persons present on Friday, 20 of whom were members of the section. Of the 53 people attending the Saturday sessions, 39 were MAA members.

The program, arranged by *A. M. Fink*, *Howard Levine* of SIAM, and *Thomas Moberg* of ASA consisted of the contributed papers, invited addresses, and the panel discussion listed below along with a program of MAA films, the Governor's report presented by *William Waltmann*, and the business session.

At the business meeting *Edward T. Hill* of Cornell College was elected to the office of Chair-person-elect while *E. James Peake* of Iowa State University was re-elected as Secretary-Treasurer. A motion was approved that the section explore the possibility of becoming involved with other nearby sections in the formation of regional seminars and conferences.

After considerable discussion two motions regarding future meetings were passed. The first, with respect to the 1982 meeting, called for a joint meeting with SIAM and ASA on a college campus and on dates that would not conflict with meeting dates of NCTM. This would preclude any joint arrangements with IAS. For the future the section decided to continue to explore the possibility of joint meetings with IAS with appropriately shared arrangements while trying to avoid conflicts with meetings of other mathematical organizations.

There were motions of appreciation for Chairman *Arnold Adelberg* and our hosts, *Charles Lindsay* and Coe College.

Papers, invited addresses, and panel discussions: "Using the Study of Consciousness in the Mathematical Classroom," *C. G. Wadsworth*, Maharishi International University; "Ancient Mathematical Models of Celestial Motion," *J. E. Mathews*, Iowa State University; "A Model for the Flow of Surfactants on the Interface of Forming Droplets," *H. A. Levine*, Iowa State University; "Teaching Operations Research at the Undergraduate Level and the Use of UMAP Materials," *A. M. Fink*, Iowa State University; "Survey Results: Responses to a Questionnaire from Mathematics Faculty of the Liberal Arts Colleges in Iowa," *D. V. Meyer*, Central College. Invited addresses: "Robust Statistical Methods," *R. V. Hogg*, University of Iowa; "Some Algorithmically Defined Functions," *R. D. Anderson*, President of MAA; "An Asymptotic Stochastic View of Split-second Hesitation," *J. C. Kegley*, Iowa State University; "On 'Optimal' Reconstruction Algorithms in Computed Tomography," *W. R. Madych*, Iowa State University. Other addresses: "Geometrical Diffraction Theory," *A.K. Cautesen*, Iowa State University; "Integration on Spheres," *K. E. Atkinson*, University of Iowa; "Identifying Comparatively Excessive Sentences of Death: A Quantitative Approach," *G. Woodworth*, University of Iowa; "The Dynamics of Electoral Competition: A Preliminary Analysis," *A. F. Kleiner* and *W. R. Collins*, Drake University; "A Statistical Analysis of the Effect of Car Inspection and Maintenance Programs on Ambient CO Concentrations," *J. Ledolter*, University of Iowa; "Terminal Skunk," *I. R. Hentzel*, Iowa State University; "Enumeration in Music Theory," *D. L. Reiner*, Grinnell College; "Genetic Polymorphism in Varying Environments," *J. L. Cornette*, Iowa State University. Panel: "Redesigning the Undergraduate Curriculum in the Mathematical Sciences--The Impact of Computers," *W. Waltmann*, Moderator, Wartburg College.

ANNUAL MEETING OF THE ILLINOIS SECTION

The sixtieth annual meeting of the Illinois SEction of the MAA convened at Illinois State University on Friday and Saturday, May 1-2, with approximately one hundred members in attendance. The meeting this year was an experimental joint meeting with the Illinois chapter of the American Association of Physics Teachers. Provost *Léon Boothe* of Illinois State welcomed the group.

Invited addresses were presented by: Professor *Bruce C. Berndt*, University of Illinois, "Editing Ramanujan's Notebooks;" *Dwight R. Diercks*, Ph.D., Argonne National Laboratory,"Coal Gasification--An Overview and Assessment;" Professor *Curtis J. Hieggelke*, Department of Mathematics and Physics, Joliet Junior College, "Interfacing Physics and Mathematics;" Professor *A. J. DiPietro*, Professor *Ferrel Atkins*, Eastern Illinois University, "A Computational Mathematics Degree Program;" Professor *Chung-Wu-Ho*, Southern Illinois University-Edwardsville, "On the Periodic Points of Continuous Functions;" Dr. *Lida K. Barrett*, Associate Provost Northern Illinois University, "Women Mathematicians--Past and Present;" Professor *David L. Quigg*, Bradley University, "UMAP Modules;" Professor *Kenneth Appel*, University of Illinois, "The Four-Color Theorem."

Professor *Maynard Thompson* of Indiana University gave the banquet address on "Some Mathematical Questions Related to Cigarette Smoking." At the Friday evening banquet, retired professor *Douglas R. Bey* of Illinois State University, longtime member of I.S.M.A.A., was presented the first annual Distinguished Service Award by the section.

At the annual business meeting Professor *John Haverhals* of Bradley University was elected Chairman for 81-82; Professor *Richard Johnsonbaugh* of Chicago State University was nominated as Chairman-elect (82-83), and Professor *John le Duc* of Eastern Illinois University was elected second Vice-Chairman.

CALENDAR OF FUTURE MEETINGS

Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.
Sixty-second Summer Meeting, Toronto, Canada, August 23–25, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Allegheny College, Meadville, Pennsylvania, April 1982.
EASTERN PENNSYLVANIA AND DELAWARE, Villanova University, Villanova, Pennsylvania, November 21, 1981.
FLORIDA, Valencia Community College, Orlando, March 5–6, 1982.
ILLINOIS, Southern Illinois University, Edwardsville, April 30–May 1, 1982.
INDIANA, Purdue University, West Lafayette, October 17, 1981.
INTERMOUNTAIN
IOWA, Grinnell College, Grinnell, March 26–27, 1982.
KANSAS, Emporia State University, Emporia, April 2–3, 1982.
KENTUCKY, University of Kentucky, Lexington, April 2–3, 1982.
LOUISIANA–MISSISSIPPI, University of Southwestern Louisiana, Lafayette, February 12–13, 1982.
MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, George Washington University, Washington, D.C., November 14–15, 1981.
METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.
MICHIGAN, first Friday and Saturday in May. Deadline for papers six weeks before meeting.
MISSOURI, University of Missouri, Rolla, April 9–10, 1982.
NEBRASKA, Kearney State College, Kearney, April 2–3, 1982.
NEW JERSEY, Trenton State College, Trenton, fall 1981.
NORTH CENTRAL, Bemidji State University, Bemidji, Minnesota, October 23–24, 1981.
NORTHEASTERN, Trinity College, Hartford, Connecticut, November 20–21, 1981.
NORTHERN CALIFORNIA, University of California, Davis, February 20, 1982.
OHIO, Lorain County Community College, Elyria, October 23–24, 1981.
OKLAHOMA–ARKANSAS, University of Arkansas, Fayetteville, March 25–27, 1982.
PACIFIC NORTHWEST, second Saturday in June. Deadline for papers six weeks before meeting.
ROCKY MOUNTAIN, Western State College, Gunnison, Colorado, April 30–May 1, 1982.
SEAWAY, SUNY, College at Brockport, Brockport, New York, November 6–7, 1981.
SOUTHEASTERN, Emory University, Atlanta, Georgia, April 1982.
SOUTHERN CALIFORNIA, University of California, Santa Barbara, November 13–14, 1981.
SOUTHWESTERN, University of Arizona, Tucson, April 1982.
TEXAS, Friday and Saturday in early April. Deadline for papers March 1.
WISCONSIN, University of Wisconsin, Fond du Lac, late March 1982.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES, Hyatt-Regency Hotel, New Orleans, Louisiana, October 7–11, 1981.
AMERICAN MATHEMATICAL SOCIETY, Cincinnati, Ohio, January 13–16, 1982.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.
ASSOCIATION FOR SYMBOLIC LOGIC, Philadelphia, Pennsylvania, December 1981.
ASSOCIATION FOR WOMEN IN MATHEMATICS
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, Canada, April 14–17, 1982.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Regency Hyatt House, Houston, Texas, October 11–14, 1981.
PI MU EPSILON
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION, Carrousel Inn, Columbus, Ohio, November 5–6, 1981.
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Netherland Hilton Hotel, Cincinnati, Ohio, October 26–28, 1981.

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

NOVEMBER

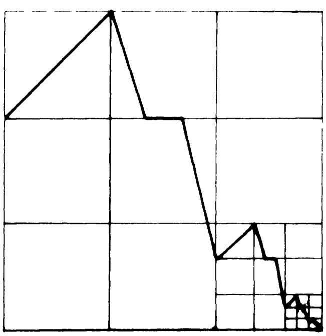
THE AMERICAN MATHEMATICAL MONTHLY

Volume 88, Number 9

Characters of Finite Groups

Normed Division Domains

Nonarchimedean Spaces



Periodic Points

Eigenvalues

Caustics

Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 **NUMBER 9**

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $14, included as part of the annual dues of $30. Students receive a 50% discount. The library subscription price is $36 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

PRINTED IN THE UNITED STATES OF AMERICA

CHARACTERS OF FINITE GROUPS: SOME USES AND MATHEMATICAL APPLICATIONS

BHAMA SRINIVASAN
Department of Mathematics, University of Illinois, Chicago Circle, Chicago, IL 60680

1. Introduction. The aim of this article is to show how the character theory of finite groups has been applied, both in group theory itself and in other areas of mathematics, to obtain interesting and deep results. Of course some choices have had to be made and many interesting examples have had to be omitted, but it is hoped that the examples mentioned here will provide a representative sample of applications of group characters.

Standard references are [**13**] for basic group theory and [**11**] for representation and character theory. A history of finite group characters has been given by Hawkins in [**7**]. An extensive survey article by Mackey [**8**] on applications of Harmonic Analysis also includes some applications of finite group characters.

A brief review of finite group characters. Let G be a finite group. A complex, or ordinary, representation of G is a homomorphism $\rho: G \to \mathrm{GL}(n,\mathbb{C})$ into the group of $n \times n$ nonsingular matrices over $\mathbb{C}$, for some n. For example, if G is a cyclic group of order m generated by a, the map $a^k \to e^{2\pi ik/m}$ gives a homomorphism of G into the group of 1×1 matrices over $\mathbb{C}$, i.e., into C^*. Another example is given by $G = A_5$, the alternating group on 5 symbols, which is the smallest nonabelian simple group and is of order 60. Since A_5 is a permutation group on 5 symbols, we can attach to each element a 5×5 permutation matrix if we think of the group as permuting a basis $\{v_1, v_2, v_3, v_4, v_5\}$ of a 5-dimensional vector space V over $\mathbb{C}$. Thus we would have, for example,

$$(12345) \to \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

and

$$(12)(34) \to \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Two representations ρ_1, ρ_2 of G are *equivalent* if there is a nonsingular matrix A such that $\rho_1(g) = A^{-1}\rho_2(g)A$ for all $g \in G$. A representation ρ is *irreducible* if it is not equivalent to a representation ρ' which is such that

$$\rho'(g) = \begin{pmatrix} \rho_1(g) & * \\ 0 & \rho_2(g) \end{pmatrix}, \quad g \in G, \tag{1.0}$$

This article is based on a talk given at a symposium in honor of Alice Schafer held at Wellesley College on April 26, 1980.

The author received her Ph.D. at the University of Manchester, England, under the direction of J. A. Green. She has held a Postdoctoral Fellowship at the University of British Columbia, has taught at the University of Madras, India, and at Clark University, and has been a Member of the Institute for Advanced Study. She is now Professor of Mathematics at the University of Illinois at Chicago Circle. Her interests are in Group Theory, especially in representation theory of finite groups and algebraic groups. She is currently the President of the Association for Women in Mathematics.—*Editors*

where $\rho_1(g)$, $\rho_2(g)$ are $l \times l$ and $(n-l) \times (n-l)$ matrices for some l, and * denotes arbitrary entries. In this case, i.e., when ρ is *not* irreducible, the maps $g \to \rho_1(g)$, $g \to \rho_2(g)$ are themselves representations of G. Thus we are primarily interested in irreducible representations of G. For example, the representation of A_5 mentioned above is not irreducible, since the element $v_1 + v_2 + v_3 + v_4 + v_5$ in V is fixed by all $g \in A_5$; and if we choose it to be the first basis element of V, we get a permutation matrix for each $g \in G$ in the form (1.0) with $l = 1$.

The function $\chi: g \to \mathrm{Trace}(\rho(g))$ of G into $\mathbb{C}$ is called the *character* of ρ. Equivalent representations have the same character. We recall that a conjugacy class of G is the set of all elements of G of the form $a^{-1}ga$, where g is a fixed element and a varies over G. The character χ of a representation is a class function on G, i.e., $\chi(a^{-1}ga) = \chi(g)$ for all $a, g \in G$. It is known that the number of equivalence classes of irreducible representations of G is equal to the number of conjugacy classes of G. Suppose the distinct irreducible characters of G are $\chi_1, \chi_2, \ldots, \chi_r$. (Here r is the number of conjugacy classes of G.) We can form a square $r \times r$ table giving the values of these characters as representatives of the conjugacy classes of G. This is called the *character table* of G. One of the characters of G is the "trivial character" which maps every element on 1; we denote this character by χ_1. For any χ_i, the integer $\chi_i(1)$ is called the degree of χ_i. We can arrange the columns so that the first column gives the degrees of the irreducible characters. The characters of an *abelian* group are all of degree 1, and in this case we have

$$\chi(g_1g_2) = \chi(g_1)\chi(g_2), \quad \text{for} \quad g_1, g_2 \in G \tag{1.1}$$

and any character χ.

We give below the character table of A_5. We note that the character of the representation given in the example, which takes the value 1 on (12)(34) and 0 on (12345), is the function $\chi_1 + \chi_2$.

	(1)	(123)	(12)(34)	(12345)	(13452)
χ_1	1	1	1	1	1
χ_2	4	1	0	-1	-1
χ_3	3	0	-1	$(1+\sqrt{5})/2$	$(1-\sqrt{5})/2$
χ_4	3	0	-1	$(1-\sqrt{5})/2$	$(1+\sqrt{5})/2$
χ_5	5	-1	1	0	0

The characters $\chi_1, \chi_2, \ldots, \chi_r$ of a group satisfy some remarkable orthogonality relations, which are as follows.

$$\frac{1}{|G|}\sum_{g\in G}\chi_i(g)\overline{\chi_j(g)} = \begin{cases} 1, & i=j \\ 0, & i\neq j. \end{cases}$$

$$\sum_{i=1}^{r}\chi_i(g)\overline{\chi_i(h)} = \begin{cases} |G|/|\mathcal{C}(g)|, & \text{if } g, h \text{ are in the same conjugacy class } \mathcal{C}(g), \\ 0, & \text{otherwise.} \end{cases} \tag{1.2}$$

This is one of the reasons that characters of finite groups are a powerful tool that we can employ in many situations.

The character table of a group G contains interesting information about G. We can determine the normal subgroups N and the characters of G/N. In particular, we can determine if G is simple or if G is solvable. If G is solvable, we can determine the Frattini subgroup (i.e., the intersection of the maximal subgroups of G). Another curious fact is that we can determine, by looking at the character table, whether or not an element is a commutator, i.e., of the form $x^{-1}y^{-1}xy$ for some $x, y \in G$. (See **[15]**, **[16]**, **[17]**, **[18]**.)

2. Number Theory. We start with a classical application of group characters in number theory. A well-known theorem of Dirichlet says that in any arithmetic progression of positive integers

$$a, a+m, a+2m, \ldots \tag{2.1}$$

where a and m are relatively prime, there is an infinite number of primes. Now, in Euler's proof of the fact that there is an infinite number of primes, he used the following identity:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1} \qquad (\text{real } s > 1).$$

Here the infinite product is over all primes, and ζ is the Riemann zeta function. From this we get

$$\begin{aligned} \log \zeta(s) &= -\sum_{p} \log(1 - p^{-s}) \\ &= \sum_{p} \sum_{k=1}^{\infty} \frac{1}{kp^{ks}}. \end{aligned}$$

Analogously, Dirichlet defined an *L-function* $L(s, \chi)$ as follows. Let Z_m^* denote the group of units, i.e., the group of elements which have inverses, in the ring Z_m of integers mod m. Let χ be a character of this finite group. Lift χ to a function on Z by

$$\chi(n) = \begin{cases} \chi(\bar{n}), & \text{if } (n, m) = 1 \\ 0, & \text{otherwise,} \end{cases} \tag{2.2}$$

where $\bar{n}$ is the image of n under the canonical map $Z \to Z_m$. Then define the function $L(s, \chi)$ by

$$\log L(s, \chi) = \sum_{p} \sum_{k=1}^{\infty} \frac{\chi(p^k)}{kp^{ks}}, \quad \text{for } s \text{ with } \operatorname{Re} s > 1. \tag{2.3}$$

We can rewrite this as

$$\log L(s, \chi) = \sum_{p} \frac{\chi(p)}{p^s} + \sum_{k=2} \frac{\chi(p^k)}{kp^{ks}}. \tag{2.4}$$

Now choose χ to be an irreducible character of Z_m^*, lifted to Z as in (2.2). Choose $b \in Z$ such that $ab \equiv 1 (\text{mod } m)$; this is possible since $(a, m) = 1$. Multiply (2.4) by $\chi(b)$ and sum over all the irreducible characters of Z_m^*. We then get

$$\sum_{\chi} \chi(b) \log L(s, \chi) = \sum_{p} \sum_{\chi} \chi(b)\chi(p) p^{-s} + \chi(b) \sum_{k=2}^{\infty} \frac{\chi(p^k)}{kp^{ks}}. \tag{2.5}$$

Now $\chi(b)\chi(p) = \chi(bp)$ (see (1.1)) and by (1.3), taking one of the elements to be 1, we get

$$\sum_{\chi} \chi(bp) = \begin{cases} |Z_m^*|, & \text{if } bp \equiv 1 \ (\text{mod}\, m), \\ 0, & \text{otherwise.} \end{cases}$$

We note that since $ab \equiv 1(\text{mod } m)$, $bp \equiv 1(\text{mod}\, m)$ is equivalent to $p \equiv a(\text{mod } m)$. Thus the first term on the right-hand side of (2.5) reduces to $|Z_m^*| \ \Sigma_{p \equiv a(\text{mod } m)} p^{-s}$. Dirichlet then shows that as $s \to 1 + 0$ along the real axis, the left-hand side of (2.5) approaches ∞ and the second term on the right-hand side approaches a finite limit. Thus $\Sigma_{p \equiv a(\text{mod } m)} p^{-s}$ is divergent, showing that there is an infinite number of primes in the arithmetic progression (2.1) (see [**4**] for the details).

Our next application takes us to Artin, who generalized the Dirichlet L-functions (see e.g., [**9**]). The Artin L-functions $L(s, \chi)$ are defined with respect to characters of certain Galois groups. Let $F \subset K$, where F, K are algebraic number fields and K is Galois over F with Galois group G. Let $\mathfrak{o} \subset F$, $O \subset K$ be the rings of integers in F and K, respectively, and suppose $\mathfrak{p}$ is a prime ideal in $\mathfrak{o}$, P a prime ideal in O, such that $\mathfrak{p} \subset P$. The *decomposition group* D_P, a subgroup of G, is defined

by $D_P = \{\sigma \in G \mid \sigma(P) = P\}$. Then we have a canonical map $D_P \to \overline{G}$, where $\overline{G}$ is the Galois group of the field O/P over $\mathfrak{o}/\mathfrak{p}$, and this map is onto with kernel I_P, the *inertia group*. Now $\mathfrak{o}/\mathfrak{p}$ is a finite field and has $N\mathfrak{p}$ elements (where N denotes the norm map). Thus there is a distinguished element in $\overline{G}$: an automorphism σ of O/P such that $\sigma(x) = x^{N\mathfrak{p}}$. This automorphism is called the Frobenius automorphism and we now regard it as an element of D_P/I_P.

Let ρ be a representation of G with character χ. We can think of ρ as a homomorphism of G into $\mathrm{GL}(V)$, the group of automorphisms of V, where V is some finite dimensional complex vector space. Let $\tilde{V}$ be the subspace of V of all elements of V fixed by I_P. Then the restriction of ρ to I_P is trivial on $\tilde{V}$ and we get a representation $\rho\colon D_P/I_P \to \mathrm{GL}(\tilde{V})$. Then the Artin L-function is defined by

$$L(s,\chi) = \prod_{\mathfrak{p}} \det\left(1 - (N\rho)^{-s}\rho(\sigma)\right)^{-1}, \qquad \operatorname{Re} s > 1$$

where the product is over the prime ideals of $\mathfrak{o}$, and this can be shown to generalize Dirichlet's definition.

Artin posed the problem of whether $L(s,\chi)$ has a meromorphic continuation to the whole plane. Brauer solved this problem using character theory, and in doing so proved a deep theorem known as "Brauer's characterization of characters," which will be described in the next section.

At this point we might also mention that Langlands has a program involving the representation theory of Lie groups and p-adic groups, which includes a vast generalization of Artin's theory of L-functions. This promises to be an exciting area of research for many years to come (see [**1**]).

3. Group Theory. We noted that characters are class functions on the finite group G. We might then ask, conversely, which class functions are characters. To pose the question in a more sensible way we have to introduce the concept of a *generalized character*. A class function $\zeta\colon G \to \mathbb{C}$ is called a generalized character (or a virtual character) if ζ can be written as an *integral* linear combination $\zeta = \sum_{i=1}^{r} n_i\chi_i$ $(n_i \in Z)$ of irreducible characters of G. The generalized characters of G form an abelian group under addition, and this group is called the *Grothendieck group* of G. For a fixed prime p, a p-elementary subgroup E of G is a subgroup of the form $E = P \times R$ where P is a p-subgroup of G and R is a cyclic group of order prime to p. We then have

(3.1) (*Brauer's characterization of characters* ([**11**], p. 81).) A class function ζ on G is a generalized character if and only if the restriction of ζ to E is a generalized character of E, for all p-elementary subgroups E (for all primes p).

To this day, this theorem remains one of the most powerful criteria we have for testing whether a class function is a generalized character. This is also an example of an oft-repeated theme in mathematics: a theorem which was conceived as a tool to prove a number-theoretic result is now a central theorem in the theory of group characters.

At this time Brauer was also developing his rich theory of modular representations and blocks with respect to a prime p. We will give a brief description of some parts of this theory and mention an application made by Brauer himself to the problem of classifying linear groups in dimension 5 (see (3.3) and the following paragraphs).

Modular representations. Let p be a prime. Let K be the cyclotomic field of $|G|$th roots of unity over the rational field Q. Then all the irreducible representations of G can be realized over K. Let P be a prime ideal of the ring of integers O of K such that P contains p. Let ρ be an irreducible representation of G. Then, in fact, the matrices $\{\rho(g)\}$ $(g \in G)$ can be taken to have entries in the local ring O_P of all elements a/b in K, where $a, b \in O$ and $b \notin P$. We can then reduce mod P to obtain a representation $\bar{\rho}$ of G over a finite field F with, say, $p^m = q$ elements, where F is isomorphic to $O/P = O_P/PO_P$. Such representations, over fields of prime characteristic, are called modular representations of the group G. We can talk about irreducible modular representations of G as in the case of ordinary representations. Brauer has defined the characters

of modular representations as complex functions on the set of p-regular elements of G (i.e., elements whose orders are prime to p) as follows. Choose an isomorphism θ of F^* into $\mathbb{C}^*$. Suppose τ is a representation of G over F, and suppose the matrix $\tau(g)$ (where g is a p-regular element of G) has the eigenvalues $\alpha^{r_1}, \alpha^{r_2}, \ldots$, where α is a generator of F^*. The Brauer character ϕ of τ is then defined by $\phi(g) = \theta(\alpha_1)^{r_1} + \theta(\alpha_2)^{r_2} + \cdots$ (we can think of this as replacing a $(q-1)$th root of unity in F^*, by a complex $(q-1)$th root of unity.). Going back to ρ and $\bar{\rho}$, we see that the character χ of ρ, when restricted to p-regular elements, is the Brauer character of $\bar{\rho}$.

Let $R_K(G)$ denote the Grothendieck group of generalized characters of G, and let $R_F(G)$ denote the analogous Grothendieck group of generalized Brauer characters of G, i.e., the free abelian group generated by the Brauer characters of the irreducible representations of G over F. Then we have a natural map $d\colon R_K(G) \to R_F(G)$, which is just the restriction to p-regular elements. Now Brauer's Theorem (3.1) has the following application (see [**11**, p. 131]).

(3.2) THEOREM. *The map d is onto, i.e., every Brauer character can be lifted to a generalized character of G.*

For example, suppose $G = \mathrm{GL}(n, q)$, the group of nonsingular matrices over F. This has a natural representation over F, where each element g is mapped onto itself. The Brauer character of this representation can be lifted to a generalized character of G. A similar result holds for the other classical groups. This has a surprising application in topology, in Quillen's proof of the Adams conjecture [**10**]. We might also mention here that Wall (see [**12**]) has used modular representations in his algebraic L-theory.

Blocks. Consider the ordinary irreducible characters $\chi_1, \chi_2, \ldots, \chi_r$ of G. We say χ_i and χ_j are in the same p-block, if there is a chain $\chi_i = \chi_{i1}, \chi_{i2}, \ldots, \chi_{im} = \chi_j$ such that if we take any two consecutive members $\chi_{il}, \chi_{i,l+1}$ in the chain and write $d(\chi_{il})$ and $d(\chi_{i,l+1})$ as sums of irreducible Brauer characters, there is some irreducible Brauer character which occurs in both. Thus the ordinary irreducible characters of G get divided into disjoint subsets, the p-blocks of G. For example, in the character table of A_5, for $p = 5$, there are two blocks; one consisting of $\{\chi_1, \chi_2, \chi_3, \chi_4\}$ and one consisting of $\{\chi_5\}$.

In particular, Brauer [**2**] developed a deep theory of p-blocks in groups where $p \mid |G|, p^2 \nmid |G|$. In such a case, let P be a Sylow p-subgroup of G, so that $|P| = p$. Assume for simplicity that $P = C(P)$ (although the theory is more general than this). Let $(N(P) : C(P)) = m$, and let $t = (p-1)/m$. ($N(P), C(P)$ denote the normalizer and centralizer of P.) Then the *principal block*, the block containing the trivial character χ_1, consists of m characters $\chi_1, \chi_2, \ldots, \chi_m$ called "nonexceptional characters" and a family of t characters all having the same degree called "exceptional characters." Let χ_{m+1} be one of these. If we form a graph with a node for each $\chi_i (i = 1, 2, \ldots, m+1)$ and join two nodes if the corresponding characters, when restricted to the p-regular elements, have an irreducible Brauer character in common, the graph turns out to be a tree. Furthermore, there are signs $b_i = \pm 1$ attached to the χ_i such that $\Sigma_{i=1}^{m+1} b_i \chi_i(1) = 0$. In the example above of $A_5 (p = 5)$, the nonexceptional characters are χ_1, χ_2 and the exceptional characters are χ_3, χ_4 in the principal block. The tree is

$$\underset{\chi_1}{\circ} - \underset{\chi_2}{\circ} - \underset{\chi_3}{\circ},$$

with signs $1, -1, 1$ attached to χ_1, χ_2, χ_3, respectively.

Brauer also showed in this case that a fragment of the character table of G has to look like a fragment of the character table of $N(P)$, with possibly some rows multiplied by -1. In our example of A_5, the portion of the table consisting of the first four rows and the last two columns is of this form. Such "local information" about the characters of G has been used extensively in constructing character tables of finite groups.

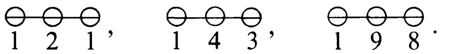
One of the first applications that Brauer made was the following (see [**2**]).

(3.3) THEOREM. *Suppose $p \mid |G|$, $p^2 \nmid |G|$, and suppose G has no normal subgroup of order p. Let n be the degree of a faithful (i.e., one-one) representation of G. Then $n \geqslant (p-1)/2$. If $n = (p-1)/2$, then $G/Z(G)$ is isomorphic to the projective special linear group* $\mathrm{PSL}(2, p)$.

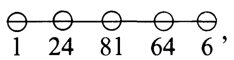
This theorem says that G cannot have a faithful representation of too small a degree, compared to the prime p, unless the Sylow p-subgroup of G is normal in G. Theorems of this nature have been proved by Feit, Leonard, Ferguson, and others (see, e.g., [**19**]).

The next application that we mention is Brauer's determination of linear groups of degree 5 (see [**3**]). By a theorem of Jordan, for any d, the group $\mathrm{GL}(d, \mathbb{C})$ has only a finite number of finite subgroups. Thus for a fixed d we can ask what finite groups have a faithful primitive irreducible representation of degree d. (To say a representation ρ is not primitive is to say that the matrices $\rho(g)$ can be taken to be monomial, where a monomial matrix is a product of a diagonal matrix and a permutation matrix; this is an easier case.) Brauer shows that, if G is a finite group with a faithful, primitive, unimodular (this means the matrices $\rho(g)$ have determinant 1) irreducible representation of degree 5, then either $G/Z(G)$ is isomorphic to the linear groups $\mathrm{PSL}(2, 11)$ or $\mathrm{PS}_p(4, 3)$ or to the alternating or symmetric groups A_5, A_6, S_5, S_6, or G is a known group of order $24 \cdot 5^4$ or a subgroup of this group. In particular, if G is simple, G is isomorphic to $\mathrm{PSL}(2, 11)$, $\mathrm{PS}_p(4, 3)$, A_5, or A_6.

We will briefly indicate how block theory enters into the proof of this theorem, taking $p = 5$. The case $\mathrm{PSL}(2, 11)$ arises by using (3.3). If we are not in this case, the order of the group is shown to be of the form $5^a \cdot 3^b \cdot 2^c$ where $a \leqslant 5$, $b \leqslant 5$, $c \leqslant 7$, and block theory is used when $a = 1$. In that case G has a Sylow p-subgroup of order 5, and the numbers m, t have the possible values $m = 4, t = 1$, or $m = 2, t = 2$. In the case when $m = t = 2$, there are 3 possible trees describing the principal block:



(The numbers give the degrees of the characters corresponding to the nodes.) We have already seen that the second tree occurs in the case of A_5, and this is indeed one of the possibilities. The third tree occurs in the case of A_6 and the first is ruled out. In the case when $m = 4$, $t = 1$, there are 11 possibilities. Most of the trees are ruled out by number-theoretic restrictions on the degrees of the characters which were developed by Brauer and Tuan. An interesting tree which occurs in this case is



which occurs in the case of the simple group $\mathrm{PS}_p(4, 3)$ of order 25920, which is also the group of the 27 lines on a cubic surface.

4. Combinatorics. Finally we mention applications of representations of finite permutation groups to Pólya enumeration theory (see [**5**], [**6**], [**14**]). Suppose D and R are finite sets and G is a permutation group on D. Let $F = R^D$, the set of all functions $f: D \to R$. We say that $f, g \in F$ are equivalent if $g = f \cdot \sigma$ for some $\sigma \in G$. Let $\mathfrak{F}$ be the set of equivalence classes, which are called *patterns*. Let $w: R \to K$ be a weight function, a map from R into a field K of characteristic 0. For $f \in F$ we define $W(f) = \prod_{i \in D} w(f(i))$. The *pattern inventory* $M(\mathfrak{F})$ is defined by $M(\mathfrak{F}) = \sum_{[f] \in \mathfrak{F}} W(f)$, where $[f]$ stands for the equivalence class of f. (Note that $W(f) = W(g)$ if f and g are equivalent.) The aim of Pólya enumeration theory is to compute $M(\mathfrak{F})$ and Pólya's theorem expresses $M(\mathfrak{F})$ in terms of the *cycle indicator* polynomial, which is defined as follows. Regarding G as a subgroup of the symmetric group S_d on d letters, where $d = |D|$, let $c_i(\sigma)$ be the number of cycles of length i in $\sigma \in G$. Then define

$$Z(x_1, x_2, \ldots, x_d) = \frac{1}{|G|} \sum_{\sigma \in G} x_1^{c_1(\sigma)} x_2^{c_2(\sigma)} \cdots x_d^{c_d(\sigma)},$$

the cycle indicator polynomial of G. Then we can state

PÓLYA'S THEOREM. $M(\mathfrak{F}) = Z(\Sigma_{y\in R}w(y), \Sigma_{y\in R}w^2(y), \ldots, \Sigma_{y\in R}w^d(y))$.

As an example, consider the following question. In how many ways can you paint the six faces of a cube in two colors, where the colorings are not distinguished by a rotation of the cube (i.e., are considered the same if one coloring can be rotated into the other)? Let G be the group of rotations of the cube, considered as a permutation group on the set D of faces of the cube. Suppose we have two colors, red and blue, and let $R = \{r, b\}$ where r, b stand for red and blue, respectively. Let $K = Q(t, u)$ where t and u are two indeterminates, and let $w : R \to K$ be the map $r \to t$, $b \to u$. Each $f\colon D \to R$ gives a way of coloring the cube and the set $\mathfrak{F}$ is the set of patterns of colorings. The cycle indicator polynomial of G is

$$\frac{1}{24}\left(x_1^6 + 3x_1^2x_2^2 + 6x_1^2x_4 + 6x_2^3 + 8x_3^2\right),$$

and thus, by Pólya's Theorem,

$$\begin{aligned} M(\mathfrak{F}) &= \frac{1}{24}\Big[(t+u)^6 + 3(t+u)^2(t^2+u^2)^2 \\ &\quad +6(t+u)^2(t^4+u^4) + 6(t^2+u^2)^3 + 8(t^3+u^3)^2\Big] \\ &= t^6 + t^5u + 2t^4u^2 + 2t^3u^3 + 2t^2u^4 + tu^5 + u^6. \end{aligned}$$

Thus, for example, there are two patterns in which four faces are colored blue and two faces are colored red, and so on.

Enumeration problems of this kind can be solved by using characters of permutation groups. An old and well-known result on permutation groups which is used in combinatorics is Burnside's Lemma [**14**]: Let P be a permutation group on a finite set X, and for $g \in P$, let $a(g)$ be the number of points of X fixed by g. Then the number of orbits of P on X is $(1/|P|)\Sigma_{g\in P}a(g)$.

Now consider P as acting on a vector space of dimension n over Q, where $n = |X|$, with basis indexed by the elements of X and P permuting the basis elements (cf. the example of A_5 in §1; the reader can easily check Burnside's Lemma there). Then $g \to a(g)$ is just the character of the representation of P obtained in this way by means of permutation matrices. In the case of Pólya's Theorem, we again have a permutation group G, and we want to evaluate a weighted sum with the function W over the G-equivalence classes in $\mathfrak{F}$. Thus we have a situation analogous to the one encountered in Burnside's Lemma; we make this precise, as follows. Let $|R| = r$, and let $D = \{1, 2, \ldots, d\}$. Let V_0 be an r-dimensional vector space over K with basis $\{e_x\}$ in one-one correspondence with the elements $\{x\}$ of R. Let V be the d-fold tensor product $V_0 \otimes V_0 \otimes \cdots \otimes V_0$. Then V is an r^d-dimensional vector space over K, and it is easy to see that we can choose a basis $\{e_f\}$ for V indexed by the elements of $f \in F = R^D$, where $e_f = e_{f(1)} \otimes e_{f(2)} \otimes \cdots \otimes e_{f(d)}$. Let G act on V by

$$\sigma \cdot e_f = e_{f(\sigma^{-1}(1))} \otimes e_{f(\sigma^{-1}(2))} \otimes \cdots \otimes e_{f(\sigma^{-1}(d))}.$$

Using the function W on R^D, we can define a linear map, also denoted W, of V by $We_f = W(f)e_f$. Then the pattern inventory $M(\mathfrak{F})$ can be shown to be $(1/|G|)\Sigma_{\sigma\in G}$ Trace $(W \cdot \sigma)$, in analogy with Burnside's Lemma. Evaluating this sum yields Pólya's Theorem.

Other enumeration problems (e.g., in graph theory) can be solved by considering a more general situation involving two permutation groups G and H. For the details the reader can consult a book by Kerber [**14**].

References

1. A. Borel, Automorphic L-functions, in Proc. Symp. Pure Math., vol. 33, American Mathematical Society, 1979.

2. R. Brauer, On groups whose order contains a prime number to the first power, I, II, Amer. J. Math., 64 (1942) 401–420, 421–440.

3. ______, Über endliche lineare Gruppen von Primzahlgrad, Math. Ann., 169 (1967) 73–96.

4. K. Chandrasekharan, Introduction to Analytic Number Theory, Springer-Verlag, 1968.

5. L. Comtet, Advanced Combinatorics, Reidel, 1974.

6. N. D. de Bruijn, Pólya's theory of counting, in Applied Combinatorial Mathematics, ed. E. F. Beckenbach, Wiley, New York, 1964.

7. T. Hawkins, Rice University Studies, vol. 64, 1978.

8. G. W. Mackey, Bull. Amer. Math. Soc., 3 (1980) 543–697.

9. J. Martinet, Character theory and Artin L-functions, in Algebraic Number Fields, ed. A. Frohlich, Academic Press, 1977.

10. D. Quillen, The Adams conjecture, Topology, 10 (1971) 67–80.

11. J. -P. Serre, Linear Representations of Finite Groups, Springer-Verlag, 1977.

12. C. T. C. Wall, Classification of Hermitian forms VI, Ann. of Math., 103 (1976) 1–80.

13. M. Hall, The Theory of Groups, Macmillian, 1959.

14. A. Kerber, Representations of Permutation Groups II, Lecture Notes in Mathematics 495, Springer-Verlag, 1975.

15. R. Brauer, Representations of finite groups, in Lectures in Modern Mathematics, ed. T. L. Seaty, Wiley, New York, 1963.

16. I. M. Isaacs, Commutators and the commutator subgroup, this MONTHLY, 84 (1977) 720–722.

17. S. Garrison, Determining the Frattini groups from the character tables, Canad. J. Math., 28 (1976).

18. G. Higman, Construction of simple groups from character tables, in Finite Simple Groups. ed. M. B. Powell and G. Higman, Academic Press, 1971.

19. P. A. Ferguson, Finite complex linear groups of degree less than $(2q+1)/3$, Proc. Symp. Pure Math., vol. 37, American Mathematical Society, 1980.

THE SET OF PERIODIC POINTS

J. -P. DELAHAYE
20 Rue Lucien Dupuis, 28500 Vernouillet, France

Introduction. The aim of this paper is to answer a question raised by T. Y. Li and J. A. Yorke in this MONTHLY [6]: "Is the closure of the periodic points [of a continuous function which maps an interval into itself] an interval or at least a finite union of intervals?"

The notion of "periodic point" may be understood in the strict sense or in a broad one, but in each case the answer is "no." This is what we show by using general methods for constructing functions with an explicitly known set of periodic points.

After some definitions we give, in Section 2, a general proposition, which is of interest in itself, concerning the set of fixed points (periodic points of period 1). In Section 3 we complete the answer when "periodic point" is understood in the strict sense, and in Section 4 we solve the other case with an appropriate counterexample.

1. Definitions. Let f be a continuous function from $[0,1]$ to $[0,1]$ (no generality is lost by choosing $[0,1]$), and let $f^p(x)$ denote the pth iterate of f.

We say that x is a periodic point of period p if:

$$f^p(x)=x \quad \text{and} \quad \forall i\in\{1,2,\dots,p-1\}: f^i(x)\neq x.$$

We say that x is an ultimately periodic point of period p if there exists some $n\in\mathbb{N}$ such that:

$$f^n(x) \text{ is a periodic point of period } p.$$

The author obtained his "Doctorat de 3ème cycle" under the direction of P. Huard. He teaches at the Lycée de Dreux and is a Research Associate at the University of Lille. His interests are in sequences in numerical analysis, sequence transformations, and acceleration of convergence, and also mathematical logic and philosophy of science. —*Editors*

We say that $x \in [0,1]$ is an asymptotically periodic point of period p if, when we define $x_0 = x$, $x_{n+1} = f(x_n)$, the sequences $\{x_{np}\}, \{x_{np+1}\}, \ldots, \{x_{np+p-1}\}$ converge to p different points.

A periodic point is an ultimately periodic point; an ultimately periodic point is an asymptotically periodic point; and a point which is not an asymptotically periodic point is called a turbulent point, because in this case the sequence $\{x_n\}$ possesses an infinite set of accumulation points.

For sets of accumulation points of such sequences see [**2**], [**4**], [**5**], [**7**], [**8**], [**10**], and for periodic points see [**1**], [**3**], [**6**], [**9**].

2. The Set of Fixed Points.

PROPOSITION. 1. (*i*) *The set of fixed points of a continuous function from* $[0,1]$ *to* $[0,1]$ *is a closed subset of* $[0,1]$.

(ii) *For every closed subset F of* $[0,1]$ *there exists a continuous function f from* $[0,1]$ *to* $[0,1]$ *whose fixed point set is F.*

Proof. (i) Trivial.

(ii) Let F be a closed subset of $[0,1]$. We assume that $0 \in F$ and $1 \in F$ (if this is not the case, it is easy to modify the proof).

The set $[0,1] - F = U$ is an open set, and consequently U is a denumerable union of disjoint open intervals (a_i, b_i), $i \in \mathbb{N}$. (We only consider the case of the infinite union.)

We define a sequence (f_n) of continuous functions:

$$\forall n \in \mathbb{N} \begin{cases} \forall x \in F \cup \left(\bigcup_{i=n}^{\infty} (a_i, b_i) \right): & f_n(x) = x \\ \forall i < n \begin{cases} \forall x \in [a_i, (a_i + b_i)/2]: \\ \forall x \in [(a_i + b_i)/2, b_i]: \end{cases} & \begin{array}{l} f_n(x) = a_i \\ f_n(x) = 2x - b_i . \end{array} \end{cases}$$

The sequence $\{f_n\}$ converges uniformly to a function f. Therefore f is continuous and one can see that its fixed point set is exactly F and that there is no other periodic point.

In Fig. 1 we show the continuous function whose fixed point set is the Cantor triadic set (i.e., the set of points of $[0,1]$ having a development in base 3 that contains only 0 and 2).

REMARKS. 1. If F is a closed set which is not a finite or denumerable union of intervals (for example, the Cantor set) the function f, above, answers the question of Li and Yorke.

2. Proposition 1 is still true when one replaces $[0,1]$ by $\mathbb{R}$ or $\mathbb{R}^n$.

3. One can also easily improve part (i) of Proposition 1 in another direction: Let f be a continuous function from $[0,1]$ to $[0,1]$ and let n be an integer different from 0. The set of periodic points of period p such that $p \mid n$ (or $p \leqslant n$) is a closed subset of $[0,1]$.

But the set of all periodic points is not necessarily closed. For example, consider the continuous function h defined as follows:

$$\forall x \in [0,1/2]: h(x) = 2x; \forall x \in [1/2,1]: h(x) = 2 - 2x.$$

One can establish (using binary developments) that the set of periodic points of h is a denumerable dense subset of $[0,1]$.

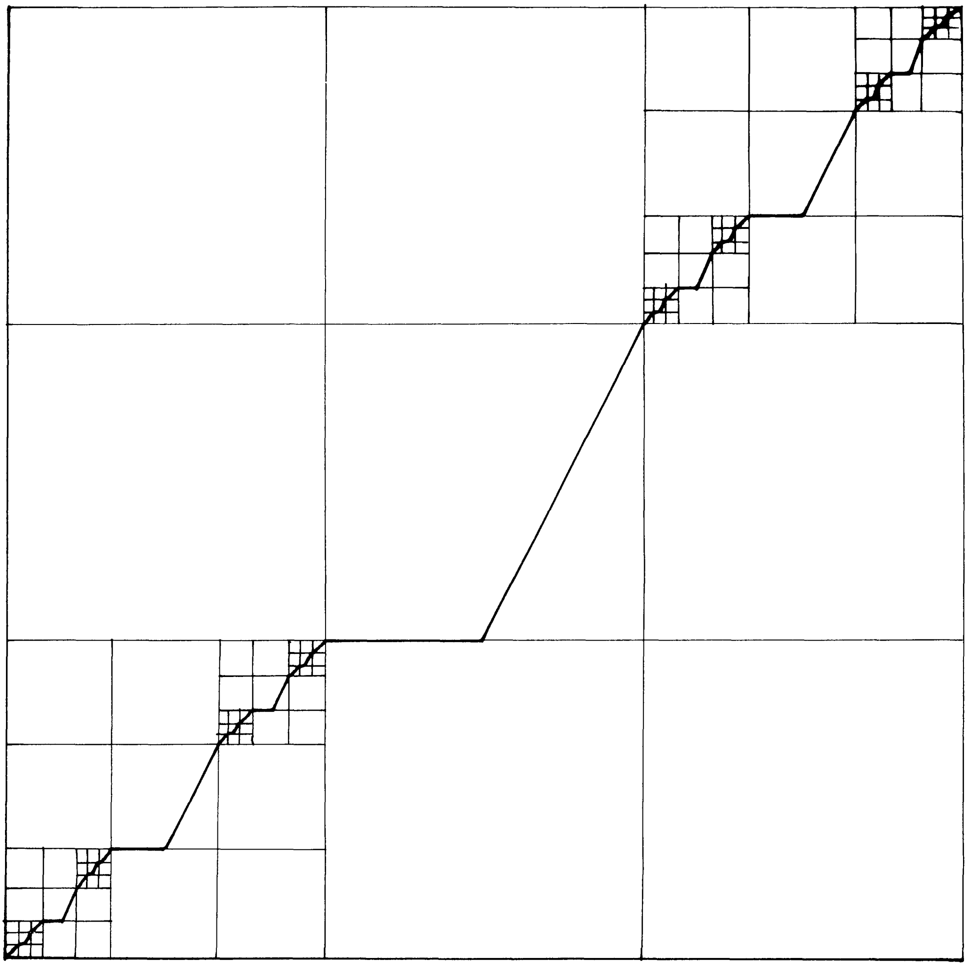
4. In [**11**] G. J. Butler and G. Pianigiani present an example of a function for which the closure of the set of periodic points is not a finite union of intervals.

3. The Set of Periodic Points. We shall say that a subset F of $[0,1]$ is symmetric if:

$$\forall x \in [0,1]: 1/2 + x \in F \Leftrightarrow 1/2 - x \in F.$$

PROPOSITION 2. (i) *The set of periodic points of period* 1 *or* 2 *of a continuous function from* $[0,1]$ *to* $[0,1]$ *is a closed subset of* $[0,1]$.

(ii) *For every symmetric closed subset of* $[0,1]$ *there exists a continuous function from* $[0,1]$ *to* $[0,1]$ *whose set of periodic points of period* 1 *or* 2 *is* $F \cup \{1/2\}$.



f

FIG. 1

Proof. (i) Trivial.

(ii) Let F be a symmetric closed subset of $[0, 1]$. We assume that $0 \in F$ and $1/2 \in F$ (if this is not the case it is easy to modify the proof).

The set $[0, 1/2] - F \cap [0, 1/2] = U$ is open, and consequently U is a denumerable union of disjoint open intervals (a_i, b_i), $i \in \mathbb{N}$. (We consider only the case of an infinite union.)

We define a sequence $\{l_n\}$ of continuous functions:

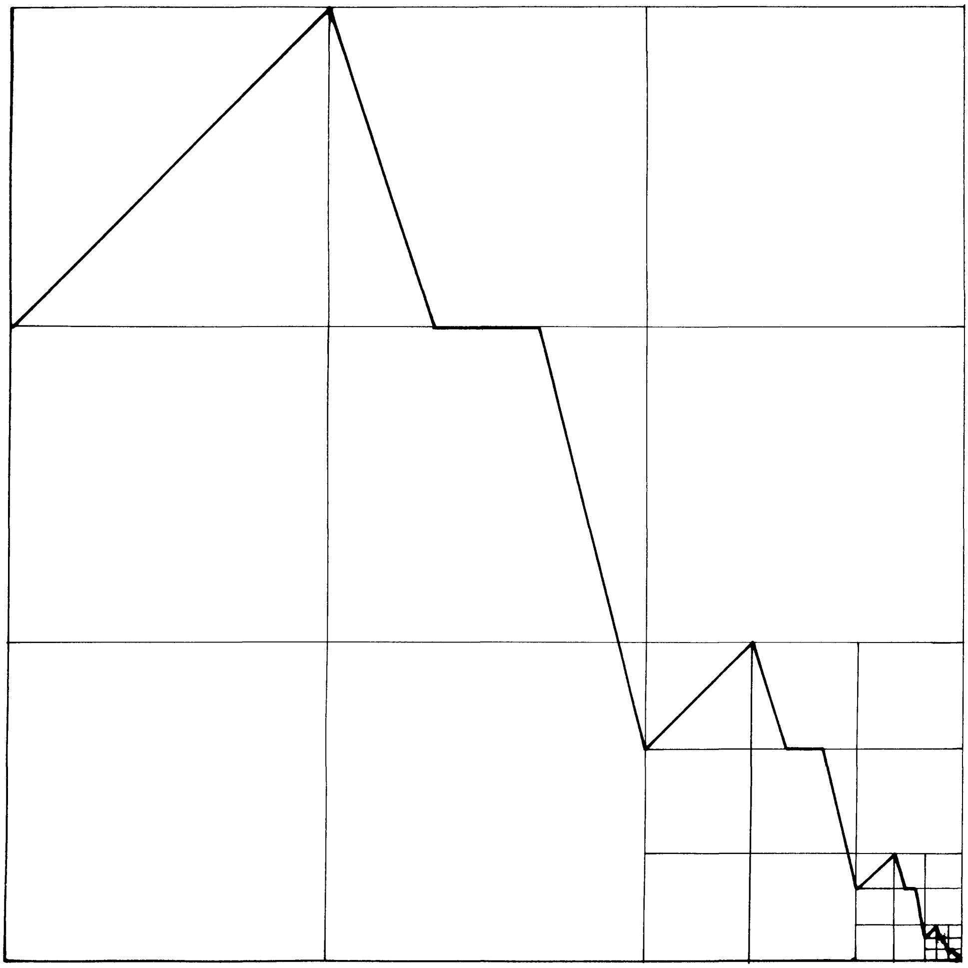
$$\forall n \in \mathbb{N} \begin{cases} \forall x \in [1/2, 1] \cup F \cup \left(\bigcup_{i=n}^{\infty} (a_i, b_i) \right): & l_n(x) = -x \\ \forall i < n \begin{cases} \forall x \in (a_i, (a_i + b_i)/2): \\ \forall x \in ((a_i + b_i)/2, b_i): \end{cases} & \begin{array}{l} l_n(x) = 1 - 2x + a_i \\ l_n(x) = 1 - b_i. \end{array} \end{cases}$$

The sequence $\{l_n\}$ converges to a function l, uniformly. Therefore l is continuous and one can see that:

—l has the unique fixed point $1/2$;

—the set of periodic points of period 2 is $F - \{1/2\}$;

—there is no other periodic point.



g

FIG. 2

REMARKS. 1. Since the Cantor set is symmetric, the function l gives us another answer to the question of Li and Yorke.

2. By analogous methods one can obtain continuous functions having only periodic points of period $2^i, 2^{i-1}, \ldots, 2, 1$ and such that the periodic point set is a very irregular closed subset of $[0, 1]$ (i.e., not a denumerable union of intervals).

4. The Set of Ultimately or Asymptotically Periodic Points. If we understand "periodic point" in a generalized sense (i.e., ultimately periodic point or asymptotically periodic point) the answer is still the same, but it is obtained in a very different way.

Let g be a function from $[0, 1]$ to $[0, 1]$. We consider the following properties:

$$g \text{ is continuous}$$
$$\forall i \in \mathbb{N} \begin{cases} \forall x \in [1 - 1/3^i, 1 - 2/3^{i+1}]: g(x) = x - 1 + 5/3^{i+1} \\ g \text{ decreases on } [1 - 2/3^{i+1}, 1 - 1/3^{i+1}]. \end{cases} \tag{*}$$

PROPOSITION 3. *If g satisfies (*), then for each $j \in \mathbb{N}$:*
(i) *$2/3^{j+1}$ is a turbulent point,*

(ii) *there exists a unique solution* y_j *of the equation*

$$g(x) = x - (3^j - 1)/3^j;$$

this point y_j *is a periodic point of period* 2^j *and*

$$y_j \in [1 - 2/3^{j+1}, 1 - 1/3^{j+1}].$$

(The proof depends on a development in base 3.)

REMARK. Functions verifying $(*)$ exist and have been used to solve another question about periodic points [3]. We now can define our last counterexample g (see Fig. 2):

$$\begin{cases} g(1) = 0 & \\ & \forall x \in [1 - 1/3^i, 1 - 2/3^{i+1}]: \quad g(x) = x - 1 + 5/3^{i+1} \\ \forall i \in \mathbb{N} & \forall x \in [1 - 2/3^{i+1}, 1 - 5/3^{i+2}]: \quad g(x) = 3 - 3x - 1/3^i \\ & \forall x \in [1 - 5/3^{i+2}, 1 - 4/3^{i+2}]: \quad g(x) = 2/3^{i+1} \\ & \forall x \in [1 - 4/3^{i+2}, 1 - 1/3^{i+1}]: \quad g(x) = 4 - 4x - 10/3^{i+2}. \end{cases}$$

The function g satisfies $(*)$; from the definition and proposition 3(i), it follows that if $x \in X = \cup_{i=0}^{\infty}[1 - 5/3^{i+2}, 1 - 4/3^{i+2}]$ then x is a turbulent point.

We have:

$$[0,1] - X = [0,4/9) \cup \left(\bigcup_{i=0}^{\infty} (1 - 4/3^{i+2}, 1 - 5/3^{i+3}) \right) \cup \{1\}.$$

For every $i \in \mathbb{N}$, $y_i \in (1 - 4/3^{i+2}, 1 - 1/3^{i+1})$. Therefore there is at least one periodic point in every interval $(1 - 4/3^{i+2}, 1 - 5/3^{i+3})$ (and consequently at least one ultimately periodic point and at least one asymptotically periodic point).

If we denote by U the set of ultimately periodic points; by A, the set of asymptotically periodic points; and by $\bar{U}$, $\bar{A}$, their closures, we have:

$$\bar{U} \subset [0,4/9] \cup \left(\bigcup_{i=0}^{\infty} [1 - 4/3^{i+2}, 1 - 5/3^{i+3}] \right) \cup \{1\}$$

$$\bar{A} \subset [0,4/9] \cup \left(\bigcup_{i=0}^{\infty} [1 - 4/3^{i+2}, 1 - 5/3^{i+3}] \right) \cup \{1\}$$

$$\forall i \in \mathbb{N} \begin{cases} \bar{U} \cap [1 - 4/3^{i+2}, 1 - 5/3^{i+3}] \neq \varnothing \\ \bar{A} \cap [1 - 4/3^{i+2}, 1 - 5/3^{i+3}] \neq \varnothing. \end{cases}$$

This implies that $\bar{U}$ and $\bar{A}$ are not finite unions of intervals.

REMARKS. 1. Is $\bar{U}$ or $\bar{A}$ a denumerable union of intervals? Even in this case (i.e., for g) the answer is unknown.

2. Like the set of periodic points, U and A are not closed in general. For example, for the function h (see Section 3) we have the surprising result that $U = A = Q \cap [0,1]$.

References

1. M. Y. Cosnard, On the behavior of successive approximations, SIAM J. Numer. Anal., 16 (1979) 300–310.
2. J.-P. Delahaye, A counterexample concerning iteratively generated sequences, J. Math. Anal. Appl., 75 (1980) 236–241.
3. ______, Cycles d'ordre 2^i et convergence cyclique de la méthode des approximations successives, Pub. A.N.O. no. 21, Université des Sciences et Techniques de Lille, 1980.

4. J. B. Diaz and F. T. Metcalf, On the set of subsequential limit points of successive approximations, Trans. Amer. Math. Soc., 135 (1968) 459–485.

5. M. Edelstein, On fixed and periodic points under contractive mapping, J. London Math. Soc., 37 (1962) 74–79.

6. T. Y. Li and J. A. Yorke, Period three implies chaos, this MONTHLY, 82 (1975) 985–992.

7. F. T. Metcalf and T. D. Rogers, The cluster set of sequences of successive approximations, J. Math. Anal. Appl., 31 (1970) 206–212.

8. G. L. Meyer and C. Raup, On the structure of cluster point sets of iteratively generated sequences, J. Optim. Theory Appl., 28 (1979) 353–362.

9. A. N. Sharkovskii, Co-existence of the cycles of a continuous mapping of the line into itself, Ukranian Math. J., 16 (1964) 61–71.

10. ______, Attracting and attracted sets, Soviet Math., 6 (1965) 268–270.

11. G. J. Butler and G. Pianigiani, Periodic points and chaotic functions in the unit interval, Bull. Austral. Math. Soc., 18 (1978) 255–265.

ON CAUSTICS OF PLANE CURVES

J. W. BRUCE
Department of Mathematics, University College, Cork, Ireland

P. J. GIBLIN
Department of Pure Mathematics, Liverpool University, Liverpool, England

C. G. GIBSON
Department of Pure Mathematics, Liverpool University, Liverpool, England

When light from a small source is reflected from the rounded inner surface of a teacup and scattered from the surface of the tea (which ideally should contain milk!) we see a bright curve on the surface. The curve will have cusps or double points which move and change when the cup is tilted.

The bright curve, or caustic, is caused by a concentration of light rays along the envelope of the reflected rays, and it only becomes visible when a screen (in this case the surface of the tea) scatters the light to our eyes.

Caustics have been studied for about 300 years, from the time of Huygens [**8**]. Old books on geometrical optics, such as [**6**] and [**7**], have something to say about them, and they appear in some books about curves, such as [**9**] and [**11**]. Cayley [**2**] wrote a memoir in 1857 in which he considered not only reflection but also refraction, and gave detailed calculations in the case of a circle.

In this paper we describe a simple geometrical technique, based on conics, for obtaining properties of caustics in the simplest case of a "mirror" which is a smooth curve in the plane, the light source also being in this plane. By this technique, various general properties, as well as special examples, can be studied directly from the mirror and without long special calculations. In

J. W. Bruce wrote his Ph.D. thesis under the supervision of C. T. C. Wall at Liverpool University. After a year at I.H.E.S., Bures-sur-Yvette, and a year as a postdoctoral fellow at Liverpool, he is now on the staff of University College, Cork, Ireland. His research interests are in singularity theory, particularly its applications to algebraic and differential geometry.

P. J. Giblin and C. G. Gibson are on the staff at Liverpool University, where they work with other members of the department in singularity theory. They have collaborated with Bruce on an extensive investigation of caustics in two or more dimensions, of which this article is the geometrical beginning. Giblin did his postgraduate work under J. E. Reeve at King's College, London; and Gibson, who was at the time primarily interested in intuitionism, under A. Heyting in Amsterdam.—*Editors*

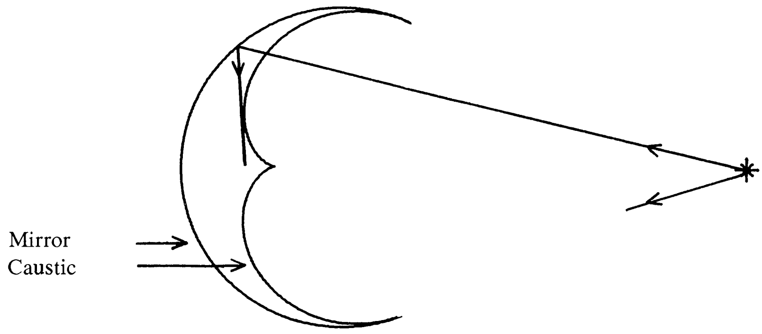


FIG. 1

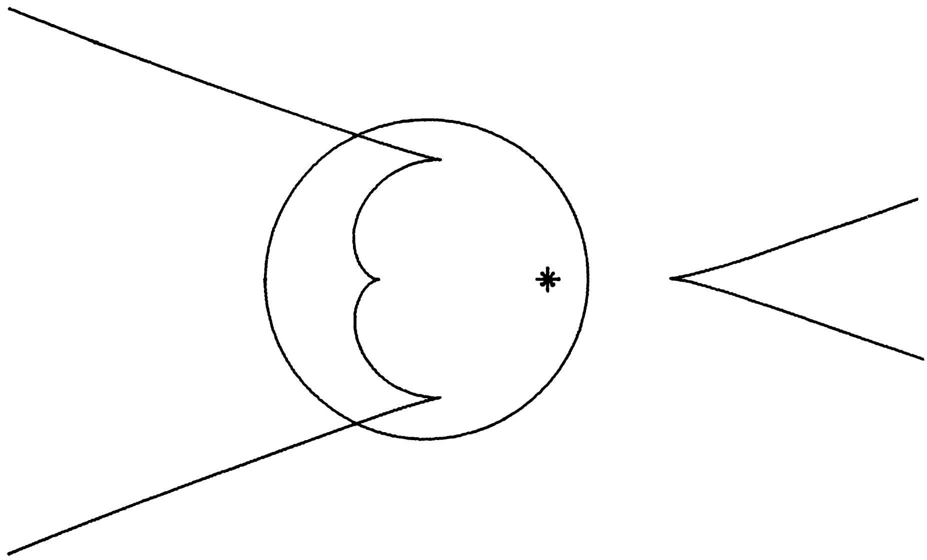


FIG. 2

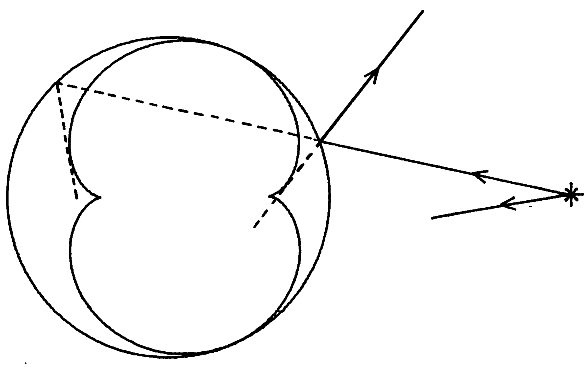


FIG. 3

a more extensive and technical article [**1**], we prove several results about the form which the caustic "generally" takes and consider caustics in higher dimensions, using, in part, techniques analogous to those described here.

In order to avoid excessive complications we consider only first reflections of light—that is, we consider the envelope of rays, coming from a single point source (which may be at infinity), after reflection from all points of the mirror. This may differ in certain respects from what is seen when a suitable screen is placed to scatter the light. For example, light rays on their way to a point of the caustic may be reflected again from another point of the mirror. We shall also assume that our mirrors have nonzero curvature; this is for a technical reason (compare (4.5)).

The teacup example is analogous to the case of a mirror forming part of a circle, and a light source outside (or maybe inside). In this case (see Fig. 1) the parts of the mathematical caustic on the same side of the mirror as the source give a good picture of what is seen; however, the caustic may also contain "virtual" parts on the wrong side of the mirror, and these will not be visible (see Fig. 2).

If we had a complete circle for the mirror and the source outside, then light would in fact only reach part of the mirror and for that part the whole caustic would be virtual (see Fig. 3). Note that the mathematical caustic in this case also includes a part formed by light rays which have penetrated the mirror once before being reflected from the concave curve beyond.

The techniques and general results are contained in Sections 1 and 2. Two special examples are described in Section 3, and in Section 4 we gather together some results on contact of curves which do not seem to be easily accessible in the literature.

We were interested to find that in a recent article in this journal [**12**], S. Schot also considered conics which have the relationship of high contact with a smooth curve (compare our (1.1) and (2.1)), and in particular he investigated the nodal cubic curve which appears in our (2.4).

1. Caustics and Conics. Suppose that M is a "mirror" in the plane, i.e., a smooth simple curve, and S a light source in the same plane (see Fig. 4). Given a light ray from S to a point P of M we can construct the reflected ray as follows: Reflect S in the tangent to M at P, giving Q, say. Then the reflected ray is QP produced; this is clear when we remember that the light reflects at the same angle to the mirror that SP makes. When P varies on M (and S remains fixed), Q describes a curve known as the *orthotomic* W of M relative to S. In fact W must be the associated "wavefront" —i.e., light emanating from all points of W, along the normals to W, at an instant, will exactly reproduce the reflected rays which result from light emanating from S at the *same* instant, and being reflected in M. Of course the fact that QP is normal to W at Q can be proved analytically—indeed it is a pleasant exercise to prove that SQ makes the same angle with the tangent to W at Q at SP makes with the tangent to M at P, and the result follows at once from this.

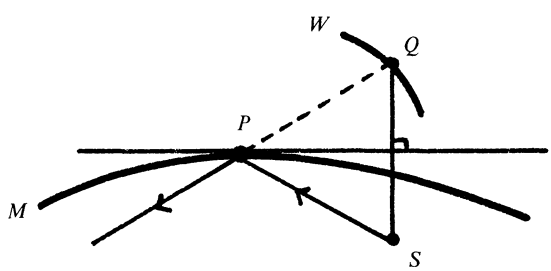


FIG. 4

The orthotomic W will be an immersed curve (free from cusps) so long as the mirror has nowhere zero curvature and the source is not on the mirror (see (4.5)). We shall always assume

that the mirror and source satisfy these conditions.

The *caustic* of M relative to S can be defined in several ways. It is, for example,

(i) the locus of centers of curvature (F, say) of W;

(ii) the envelope of the reflected rays, i.e., the envelope of normals to W. Either of these curves is known as the *evolute* of W (see, e.g., [**4**], [**7**], [**10**]).

It is worth noting that in view of (ii) the tangent to the caustic at F will pass through the corresponding point P on M.

One of the aims of this paper is to describe the caustic in terms of M and S alone, eliminating the intermediary W.

In view of (i) we shall be particularly interested in the circle of curvature of W at Q: the unique circle having at least 3-point contact with W at Q. (See Section 4 for details of n-point contact.) If the curvature of W at Q is zero, this circle becomes the tangent to W at Q, and the point F on the caustic is at infinity. Let us ask the question: Which curve will have this circle of curvature as its orthotomic relative to S?

PROPOSITION (1.1). *Suppose that the curvature of M at P is nonzero and that the tangent to M at P does not pass through S. Then there is a unique nonsingular conic, with S as one focus, having at least 3-point contact with M at P. When W has nonzero curvature at Q, the orthotomic of this conic relative to S is the circle of curvature of W at Q and the conic is an ellipse or hyperbola according as S lies inside or outside the circle.*

When W has zero curvature at Q, the conic is a parabola whose orthotomic relative to S is the tangent to W at Q.

The other focus of the conic (when the conic is a parabola, this means the point at infinity on the axis) is the center of curvature of W at Q, i.e., the point of the caustic corresponding to P.

Notes. Note the immediate consequence that, for fixed M, if the caustic relative to S passes through F, then the caustic relative to F passes through S (see (1.4) for the case F at infinity).

We shall in fact give two proofs of the first statement of (1.1), one algebraic and one geometric. The algebraic details are needed later, in (1.5) and (2.4).

In fact the first statement of (1.1) can also be deduced from the following easily proved theorem of complex projective geometry. Given three distinct lines l, m, n and a point P of n such that l, m do not meet on n and do not pass through P, then there is a *unique* irreducible conic touching n at P, having assigned "curvature" there, and touching l and m. We then take n to be the tangent to M at P and l, m the lines from S to the circular points at infinity to deduce the required result. The *reality* of the conic follows from uniqueness: the given data are invariant under conjugation; so the conic will be also.

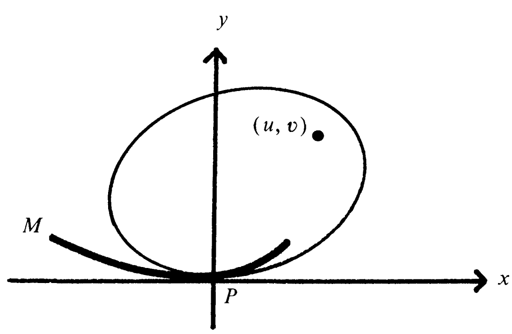


FIG. 5

Proof of (1.1). Take P as the origin $(0, 0)$ and the tangent to M at P as the line $y = 0$ (see Fig. 5). Then any nonsingular conic touching M at P has the form

$$ax^2 + 2hxy + by^2 + y = 0 \tag{1}$$

where $a \neq 0$. The curvature of this conic at $(0,0)$ is $y''(0) = -2a$; so for 3-point contact or more with M we have $-2a =$ curvature of M at P, and hence a is fixed and nonzero. (Compare (4.1).) A standard calculation (see, e.g., [**13**], p. 140) shows that the necessary and sufficient conditions for (u, v) to be a focus of (1) are

$$4(h^2 - ab)(v^2 - u^2) = 4av + 4hu + 1 \tag{2}$$

and

$$2(h^2 - ab)uv = au - hv. \tag{3}$$

We now suppose a, u, and v are given (a and v nonzero) and seek b and h. (We can take $v \neq 0$ since S is not on the line $y = 0$.) Now $(2)\frac{1}{2}uv - (3)(v^2 - u^2)$ gives a linear equation for h in which the coefficient of h is $v(u^2 + v^2) \neq 0$. Hence h is unique, and real. Also $v^2 - u^2$ and uv cannot both be zero, and so (2) or (3) then gives b uniquely.

The geometric argument is as follows. It is clear from the construction of W from M that we can recover M from W by taking the envelope of perpendicular bisectors of lines SQ for Q on W. (M is the *antiorthotomic* of W relative to S.)

When W has nonzero curvature at Q, we have to show that, replacing W by its circle of curvature at Q, the corresponding antiorthotomic is an ellipse or hyperbola with one focus at S and the other at the center of curvature, F, say. The locus of midpoints of SQ', where Q' varies on the circle of curvature, will be another circle R whose center is the midpoint of SF. We have to find the envelope of lines through points X of R perpendicular to SX (see Fig. 6): this is the antipedal or negative pedal of R relative to S (compare [**11**], p. 157). Taking R to be $x^2 + y^2 = c^2$ and S to be $(d, 0)$ it is a pleasant exercise to verify that the negative pedal is an ellipse or hyperbola, according as S is inside or outside R, with one focus at S and center at the center of R (indeed, R is the auxiliary circle of the conic). The other focus, F', say, is therefore such that the midpoint of SF' is the center of R; hence $F' = F$, as required. Clearly S is inside or outside R according as S is inside or outside the circle of curvature of W at Q. The point S can only lie *on* the circle of curvature (equivalently, on R) if S is on the tangent to M at P, which is excluded by hypothesis. (See (4.2).) Finally, the conic, whose orthotomic has been shown to be the circle of curvature, has at least 3-point contact with M at P because orders of contact are preserved (see (4.4)).

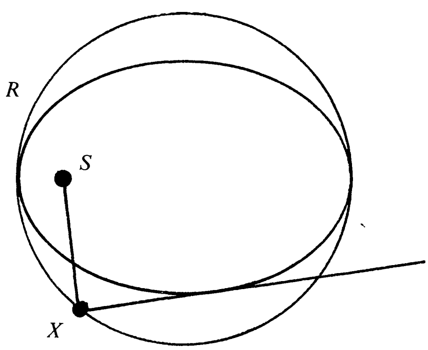


FIG. 6

The case when the curvature of W at Q is zero presents no additional difficulty and will be left to the reader. □

The primary importance of (1.1) is that we can work directly with M when studying the caustic, rather than having to go via W.

The two (real) foci of an ellipse are on the same side of any tangent, while those of a hyperbola are on opposite sides. Thus the point F of the caustic corresponding to P is on the right side of M to be visible if, and only if, the conic with at least 3-point contact is an ellipse. When the conic is a hyperbola, we get "virtual" points of the caustic—on the wrong side of the mirror.

In (1.5) we shall find a condition on M, rather than W, which determines the nature of the conic with at least 3-point contact.

(1.2) CONSTRUCTION OF A POINT ON THE CAUSTIC. Suppose we are given S, P and the center of curvature C of M at P. These ingredients uniquely determine the conic of (1.1); is there some simple construction of the other focus F, i.e., the corresponding point on the caustic? The conic will also have center of curvature C at P (having the same normal and curvature—see (4.1)). There is a well-known construction (see, e.g., [**14**, p. 95], [**15**]) for the center of curvature of a conic at P given one focus S, the normal at P, and the axis through S: In Fig. 7, C is the center of curvature at P. Given P, S, and C, we can work the construction backwards to give the axis. This gives one line on which the other focus F lies. Another line is given by taking an angle on the other side of PC, equal to angle SPC. This line meets the axis on one side of S for an ellipse, at infinity for a parabola, and on the other side of S for a hyperbola.

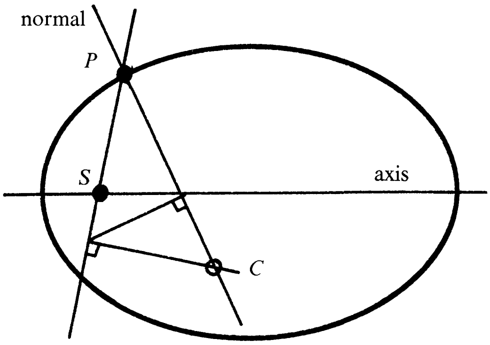


FIG. 7

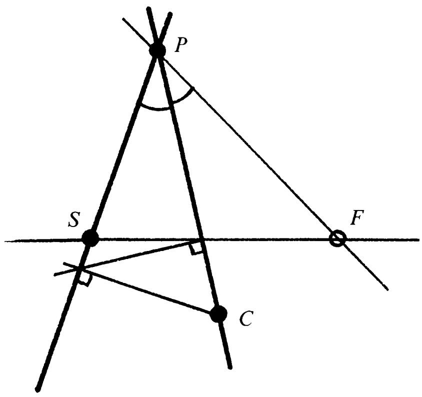


FIG. 8

Hence, given S, P, and C, we can construct F as in Fig. 8.

Of course this construction breaks down when S lies on PC. Then PC is the axis of the conic containing the (real) foci, and the other focus is on the same line. Let $PS = d$ and $PC = r$. Then easy calculations show the following.

(i) If S, C are on the same side of P, and $2d < r$, then F is on the opposite side of P, at a distance $dr/(r - 2d)$ from P (and the conic is a hyperbola);

(ii) If S, C are on the same side of P, and $2d > r$, then F is on the same side of P, at a distance $dr/(2d - r)$ from P (and the conic is an ellipse);

(iii) If S, C are on opposite sides of P, then F is on the same side of P as C, at a distance $dr/(2d + r)$ from P (and the conic is a hyperbola).

(iv) In case (i) or (ii), with $2d = r$, we get a parabola.

(1.3) TANGENT AT P PASSES THROUGH THE SOURCE S. There are two hypotheses on M in (1.1). The first (curvature nonzero) is necessary in order to avoid singularities on the orthotomic W (compare (4.5)).

The second (S not on the tangent to M at P) is only necessary in order to allow the conic of (1.1) to be nonsingular and unique. When S lies on the tangent at P, but not at P itself, the orthotomic W passes through S and the normal to W is the line SP (see Fig. 9), so that the center

of curvature of W at S lies on SP. In fact a direct calculation (compare (4.2)) shows that the center of curvature is actually P, and that *the caustic touches M at P*. Finally, it follows from (1.1) that the point of the caustic corresponding to P can never be at P when S is not on the tangent to M at P, since a focus at a finite point cannot lie on a nonsingular conic.

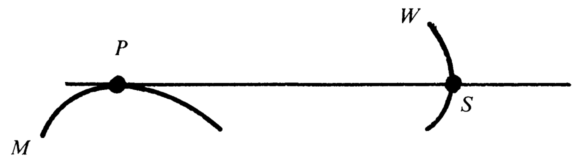


FIG. 9

If there are other tangents to M from S, then W will have more than one branch at S; but these branches will clearly have distinct tangents at S so long as there are no double tangents to M from S.

It is natural to ask what happens when the light source S is at infinity, i.e., when the incident light is parallel to a given direction. The orthotomic as defined above no longer makes sense, but we can redefine it as follows. Let I be a straight line perpendicular to the incident light (so I is the "incident wavefront"). For any point A of I, find its reflection Q in the tangent to M at the point P where the ray through A meets M (see Fig. 10). Then the locus of Q, as A varies on I, is the orthotomic. Of course we get different orthotomics for different lines parallel to I, but these will be parallel curves (i.e., curves having the same normals). As I advances to the left in the diagram, the point Q moves down the straight line QP, which is normal to all the different orthotomics. These parallel curves will therefore have the same envelope of normals; so, using definition (ii) of the caustic, the latter is well defined.

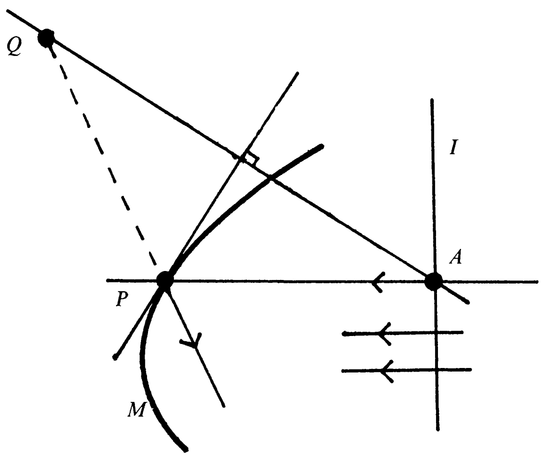


FIG. 10

By regarding M as a curve in the real projective plane and the light source S as a point at infinity, we can reformulate (1.1) so as to apply to this situation. The result is simply a limiting case of (1.1), as follows.

(1.4) LIGHT SOURCE AT INFINITY. *Suppose that the curvature of M at P is nonzero, and that the tangent to M at P does not pass through S (i.e., in the finite plane the tangent is not parallel to the*

incident light). Then there is a unique parabola touching the line at infinity at S (i.e., having axis parallel to the incident light) and having 3-point contact or more with M at P. The focus of this parabola is the point of the caustic corresponding to P. □

In the case when the tangent to M at P passes through S (i.e., is parallel to the incident light) we find again that the caustic touches M at P. (For an example of this, see (3.2).)

It is interesting to find the condition on M and P under which the unique conic of (1.1) is an ellipse, a parabola, or a hyperbola. Let us define the *discriminant circle* of M at P to be the circle touching M at P, whose center is $\frac{1}{4}$ of the way from P to the center of curvature, C, of M at P.

(1.5) PROPOSITION. *With the hypotheses of* (1.1), *suppose that S and C lie on the same side of the tangent at P to M. Then the unique conic of* (1.1) *is an ellipse, a parabola, or a hyperbola according as S lies outside, on, or inside the discriminant circle.*

When S and C lie on opposite sides of the tangent, the conic is always a hyperbola.

We can think of this as saying (in the first case) that when S is "too close to P" the rays of light are diverging too strongly to start converging after reflection from M.

Proof. It follows from equations (2) and (3) in the proof of (1.1) that, in the notation used there,

$$h^2 - ab = \frac{4a(u^2 + v^2) + v}{4v(u^2 + v^2)}.$$

Now the conic is an ellipse, a parabola, or a hyperbola according as $h^2 - ab$ is < 0, $= 0$, or > 0. We can take $v > 0$, and then the point (u, v) is on the same side of $y = 0$ as the center of curvature if, and only if, $a < 0$. Since the locus of (u, v) for which $h^2 = ab$ is a circle, center $(0, -1/8a)$, radius $1/8a$, and the center of curvature is at $(0, -1/2a)$, the result follows. □

2. Singularities of Caustics. Since the caustic of M relative to the light source S is the evolute of another curve, W, singularities of the caustic will correspond to the points of W where the derivative of the curvature is zero, i.e., to the *vertices* of W. (Compare [**4**], [**5**], [**10**].) These will give *cusps* on the caustic; we consider double points (nodes) later in this section. At a point Q of W where the curvature is stationary, the circle of curvature of W at Q has the same tangent, curvature, and derivative of curvature as W, since a circle has constant curvature; and hence the circle has at least 4-point contact with W at Q (see (4.1)). It follows from (4.4) that the conic of (1.1) has at least 4-point contact with M at P. The converse works in the same way; hence we have the following result.

(2.1) PROPOSITION. *With the hypotheses of* (1.1), *the point of the caustic corresponding to P on M is a cusp if, and only if, the unique conic having at least* 3-*point contact with M at P, and having S as one focus, actually has at least* 4-*point contact.* □

(2.2) REMARKS (1) Higher cusps on the caustic correspond to higher vertices of W, i.e., to points at which the first and second derivatives of curvature (say with respect to arc length) vanish. At such points the circle of curvature will have at least 5-point contact with W, and as a result the conic of (1.1) will have at least 5-point contact with M. Note that at a given point of M (with curvature nonzero) there will be a *unique* nonsingular conic having 5-point contact, hence only two possible positions for S. The computer drawing (see Fig. 11) shows the locus of foci of 5-point contact conics (which we called the *focoid* of M in [**1**, §5]), for the case when M is the cubic oval $y^2 = 0.64(x - x^3)$, $0 \leqslant x \leqslant 1$. Thus *for S off the focoid, the caustic of M will have no higher cusps*. Note that the focoid in the drawing is a connected curve (except for one cusp at infinity), doubly covering the oval. Also the focoid itself has cusps: these are the foci of the 6-point contact conics, which touch M at the sextactic points of M.

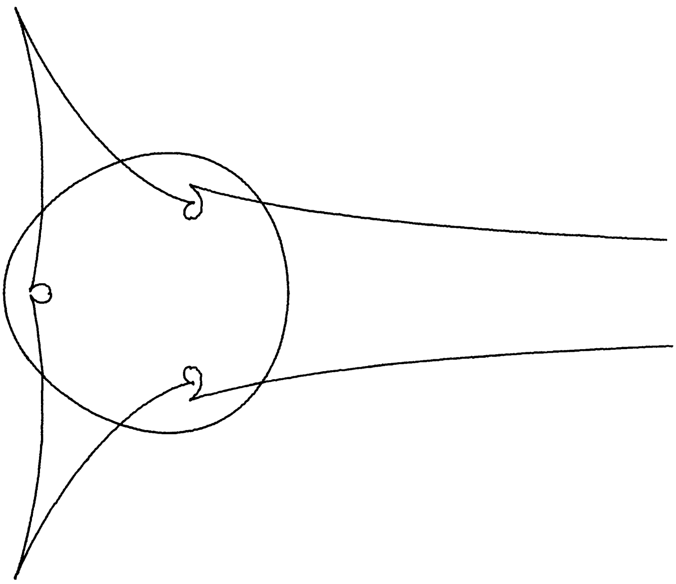


FIG. 11

(2) It follows at once from (2.1) that, if F is a cusp when the source is at S, then (with the same M) S is a cusp when the light source is at F.

(3) With the curvature of M at P nonzero, but S on the tangent to M at P (not at P itself), we find that the corresponding point of the caustic is never a cusp. (This follows on putting $p = 0$ in the formula for $d\kappa(W)/dt$ in the proof of (4.4).)

(4) The cuspidal tangent will always pass through the corresponding point P of M. (The tangent at an ordinary point of the caustic will pass through the corresponding point of M, by definition (ii) of the caustic in Section 1. This is a limiting case of the latter result.)

(5) With S at infinity (parallel incident rays), we find that the unique parabola with axis in the direction of the light and having at least 3-point contact with M at P actually has at least 4-point contact precisely when P correspond to a cusp of the caustic.

(6) In the case when M is itself an ellipse, and S one focus, the unique conic of (1.1) is always M itself, and all points of the caustic are *very* singular: the caustic collapses to the other focus of M.

As an illustration of these ideas we prove the following.

(2.3) PROPOSITION. *Suppose that the normal to M at P passes through S. Then P gives a cusp on the caustic if, and only if, P is a vertex of M.*

Proof. Clearly the point on the caustic corresponding to P also lies on the normal; so this normal is the axis of the conic of (1.1) joining the foci (see Fig. 12). In that case P is a vertex of the conic, which will therefore have at least 4-point contact with M at P if and only if P is a vertex of M. $\square$

Given a point P of M we can find the locus of points S which, as light source, make P correspond to a cusp on the corresponding caustic. By (2.1) this is the locus of foci of conics having at least 4-point contact with M at P. (This locus was also discussed by Schot in [**12**].)

(2.4) PROPOSITION. *Provided P is not a vertex of M, the locus above is a nodal cubic curve with the node at P and nodal tangents along the tangent and normal to M at P. The nodal cubic breaks up into a circle and the diameter through P in the special case when P is a vertex of M. The circle is the one touching M at P and passing through the center of curvature of M at P.*

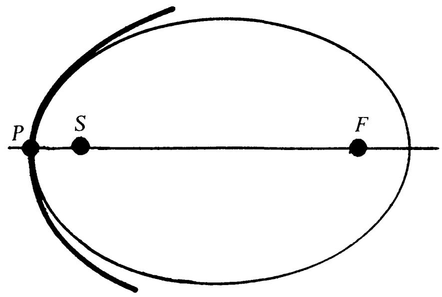


FIG. 12

Proof. We use the notation of the proof of (1.1). If the conic (1) is to have at least 4-point contact with M at $(0,0)$ then not only $y''(0)$ but also $y'''(0)$ is given. Now $y''(0) = -2a \neq 0$ and $y'''(0) = 12ah$; so in effect a and h are given and the only parameter is b. (That is, the conics with at least 4-point contact form a *pencil*. They also form a dual pencil, or *range*, but we do not use this fact here.) Eliminating b from (2) and (3) we get

$$2(v^2 - u^2)(au - hv) = uv(4av + 4hu + 1).$$

Replacing (u, v) by (x, y) and rearranging, we get as the locus of foci

$$2(x^2 + y^2)(ax + hy) = -xy. \tag{4}$$

(Note that $ax + hy = 0$ is the "axis of aberrancy" [**12**], i.e., the locus of centers of the conics.)

The cubic curve (4) always has a node at $(0,0)$ with nodal tangents the x and y axes. If it breaks up, then one component must be a line through $(0,0)$. But using $a \neq 0$ the only possible such line is $x = 0$, and this can only be a component when $h = 0$; i.e., $y'''(0) = 0$. But this is also the condition for $(0,0)$ to be a vertex of M. The assertion about the residual circle is easy to check from (4), using the fact that $(0, -1/2a)$ is the center of curvature. □

Note that the cubic is "circular"; i.e., in the complex projective plane it passes through the circular points at infinity. Indeed, that is why the residual conic above must be a circle. Hence there is just one real point at infinity on the cubic (corresponding to the unique parabola with at least 4-point contact) and therefore exactly one direction for parallel incident light which makes P correspond to a cusp on the caustic. The cubic also contains all the foci of the conics with at least

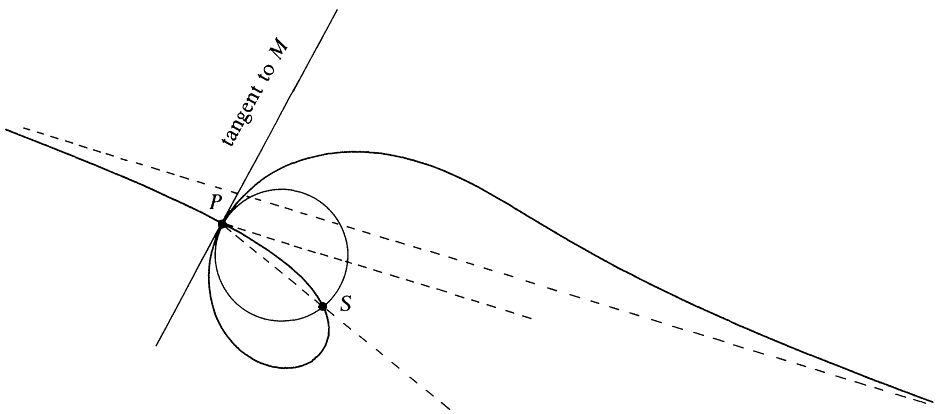


FIG. 13

4-point contact, including the complex ones. Given one point S of the real cubic curve, the other real focus of the same conic is easily found from the fact that the normal to M at P bisects the angle SPF. In particular, if we take the intersection (away from P) of the cubic and the discriminant circle of (1.5), this will give us the focus of the unique parabola having at least 4-point contact, and the bisector construction will give the axis of that parabola, which is parallel to the real asymptote of the cubic (see Fig. 13).

As an example, take M to be an ellipse. It turns out that the nodal cubic of (2.4) always has the following properties:

(i) It passes through the foci of M, and through the perpendiculars from P to the principal axes.

(ii) It intersects M in precisely those points of M at which the normal to M passes through P.

Fig. 14 is a computer drawing of the cubic for two vertices of an ellipse and one intermediate point, marked P. By examining how the cubic varies as P moves round the ellipse, it is possible to prove that any interior point S of the ellipse, other than a focus, lies on the cubic curve corresponding to precisely four positions of P. For example, when S is on a principal axis (not at a focus) of M, the four positions are the ends of that axis and the ends of the chord through S perpendicular to that axis. Hence there are always exactly four cusps on the caustic of an ellipse, with S inside and not at a focus.

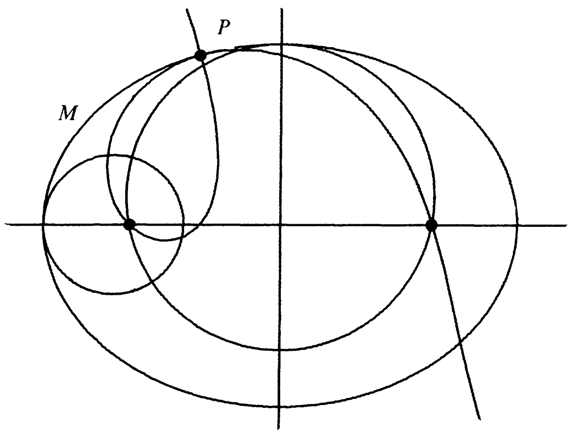


FIG. 14

There is another kind of singularity which a caustic can have, namely, a *double point* or *node*, where two nonsingular branches cross. Usually the tangents to the branches will be distinct. Double points are much harder to detect than cusps, since they depend on properties of M close to two points instead of one. Using (1.1) it is easy to see the following.

(2.5) PROPOSITION. *With the hypotheses of* (1.1), *two distinct points P, P' of M will give a double point on the caustic if, and only if, there exist two confocal conics, one of the common foci being at S, having* 3*-point contact with M at P and P', respectively. The other common focus is the location of the double point, and the nodal tangents pass through P and P', respectively.* (*Of course in special cases the two confocal conics might coincide.*) □

Using this criterion the reader can probably convince himself that the caustics of a circle (compare Cayley [**2**] and (3.1) below) never have a double point. For it is not hard to convince oneself that two confocal conics could never have the same circle of curvature M at distinct points P and P'. (In this case it is clear that the conics could not coincide, for we should then have a conic and a circle with six common points.)

The "global" questions posed by double (or triple!) points are on a more technical level than the "local" questions considered here (compare [**1**]), and we content ourselves with an example, (3.2) in the next section.

3. Examples.

(3.1) THE CIRCLE. The case when M is a circle was studied in detail by Cayley [**2**] (and even earlier by others—see Cayley's references), but let us attempt to cast further light on the subject by the use of conics rather than complicated special calculations. We shall take S inside the circle.

Now every point of the circle is a vertex (the curvature being constant), so the cubic curve of (2.4) breaks up into a circle of half the radius of M, together with the diameter through P. Given a point S inside M, which points P of M have the property that the associated cubic curve passes through S? A little thought (but no reflection) shows that there are four such points: the ends of the diameter through S, and two other points, symmetrical about this diameter. One of these is shown in Fig. 15; since the angle in a semicircle is a right angle, SP is at right-angles to the diameter SC. Furthermore the cusp F corresponding to P will be at the reflection of S in PC (since PC will be the minor axis of the 3-point contact conic). Hence two cusps will be on the circle, center C, through S. The cuspidal tangents pass through the corresponding points P—see (2.2) (4). (Cayley, by calculating the equation of the caustic, observes these facts too. The caustic is in fact the famous sextic with four real cusps and two others at the circular points at infinity.)

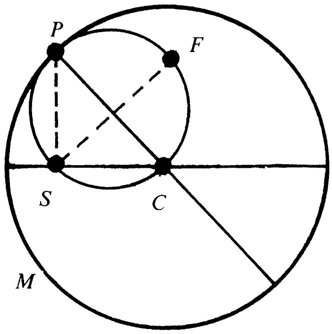


FIG. 15

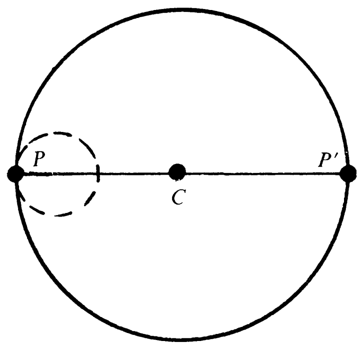


FIG. 16

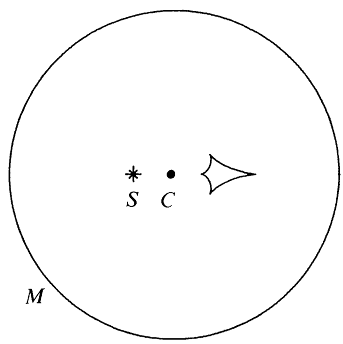


FIG. 17

The two cusps, corresponding to the ends P and P' of the diameter through S, will also lie on this diameter. Consider the discriminant circle of (1.5), drawn dotted in Fig. 16. This meets the diameter halfway between P and C. For S inside the discriminant circle, the conic with at least 3-point contact is a hyperbola and the cusp corresponding to P is outside M to the left; for S on the discriminant circle the conic is a parabola and the cusp at infinity; for S outside the discriminant circle the conic is an ellipse and (for S to the left of C) the cusp is to the right of S. The distances are given by (i), (ii), (iii) of (1.2).

As an example, when the source is a quarter of the way from C to the circumference, we get the picture in Fig. 17.

(3.2) PARABOLA, LIGHT SOURCE AT INFINITY. This example is also rich in geometry, and we shall indicate only the results, leaving the reader to fill in the details. But first let us see, using (2.2) (3), why there can be *no cusps* on the caustic. For when M is a parabola (and S at infinity), any parabola having at least 4-point contact with M must be M itself. (In projective terms, M and such a parabola will touch the same five lines, one being at infinity, and so will coincide.) But that implies that a cusp can only arise when the light is parallel to the axis of M, in which case the caustic collapses to the focus of M.

It is instructive to carry out the actual calculation of the caustic, which can be done directly from (ii) in Section 1, or by finding the foci of the conics with at least 3-point contact, as in (1.4). Let the parabola M be $y^2 = 4ax$ $(a > 0)$ and suppose the incident light is parallel to the line $x = by$. The point on the caustic corresponding to $(at^2, 2at)$ on M turns out to be

$$(x, y) = \frac{a}{1 + b^2}\left(bt^3 + 3t^2 - 3bt + b^2, -t^3 + 3t^2b + 3t - b\right),$$

so that the caustic is a cubic curve. Note that

$$by + x = 3at^2$$

and

$$y - bx = -at^3 + 3at - ab.$$

Using the last two equations it is easy to check that t and t', with $t < t'$, give the same point on the caustic if, and only if, $t = -\sqrt{3}$ and $t' = \sqrt{3}$; hence there is a *node* at $(a(9 + b^2)/(1 + b^2),$ $8ab/(1 + b^2))$, which, for all b (and fixed a, i.e., the same M) lies on a circle, center $(5a, 0)$ and radius $4a$. The locus of nodes, for varying incident directions, is therefore a circle through the focus $(a, 0)$ of M. This circle is tangent to the parabola at precisely the points $t = \pm\sqrt{3}$, and of course the nodal tangents must pass through these points. It is amusing to note that the light ray through $(9a, 0)$, after reflection from the x-axis, cuts the circle again precisely at the node.

The caustic is a nodal cubic curve which has no asymptotes (it inflects the line at infinity). The point given by $t = 0$ lies on the line $by + x = 0$, and since the tangent to the caustic here must pass through the point $t = 0$ on the parabola, i.e., through $(0, 0)$, the caustic touches this line. Fig. 18 shows the case $b > \sqrt{3}$, i.e., angle of incidence with the x-axis less than 30°.

4. Appendix on Orders of Contact. We collect together here some results on orders of contact which are used in the preceding sections. We are concerned with smooth parametrized curves in the plane $\mathbb{R}^2$.

Suppose that two curves have a common tangent at a common point. Taking the point as $(0, 0)$ and the tangent as $y = 0$, we can parametrize the curves locally by x, writing them as $y = f(x)$, $y = g(x)$, where f and g are smooth, and $f(0) = f'(0) = g(0) = g'(0) = 0$.

DEFINITION. The curves above have *n-point contact* at $(0, 0)$ if

$$f^{(k)}(0) = g^{(k)}(0) \qquad \text{for} \quad k = 2, 3, \ldots, n - 1$$

while

$$f^{(n)}(0) \neq g^{(n)}(0),$$

where the derivatives are with respect to x.

The curvature of the curve $y = f(x)$ at $(x, f(x))$ near $(0, 0)$ is

$$\kappa(x) = f''(x)\left(1 + (f'(x))^2\right)^{-3/2}.$$

Hence $\kappa(0) = f''(0)$, and by differentiation of $\kappa(x)$, $\kappa'(0) = f'''(0)$. Also, if s denotes arc length along the curve, and dots denote differentiation with respect to s, then

$$\dot{\kappa} = \kappa'\dot{x} = \kappa'\left(1 + (f'(x))^2\right)^{-1/2}.$$

Hence $\dot{\kappa}(0) = \kappa'(0)$. This proves the following.

(4.1) PROPOSITION. *Two smooth curves having a common tangent at a common point P have*
(i) *at least 3-point contact at P if, and only if, they have the same curvature at P;*
(ii) *at least 4-point contact at P if, and only if, they have the same curvature and the same first derivative of curvature with respect to arc length, at P.* □

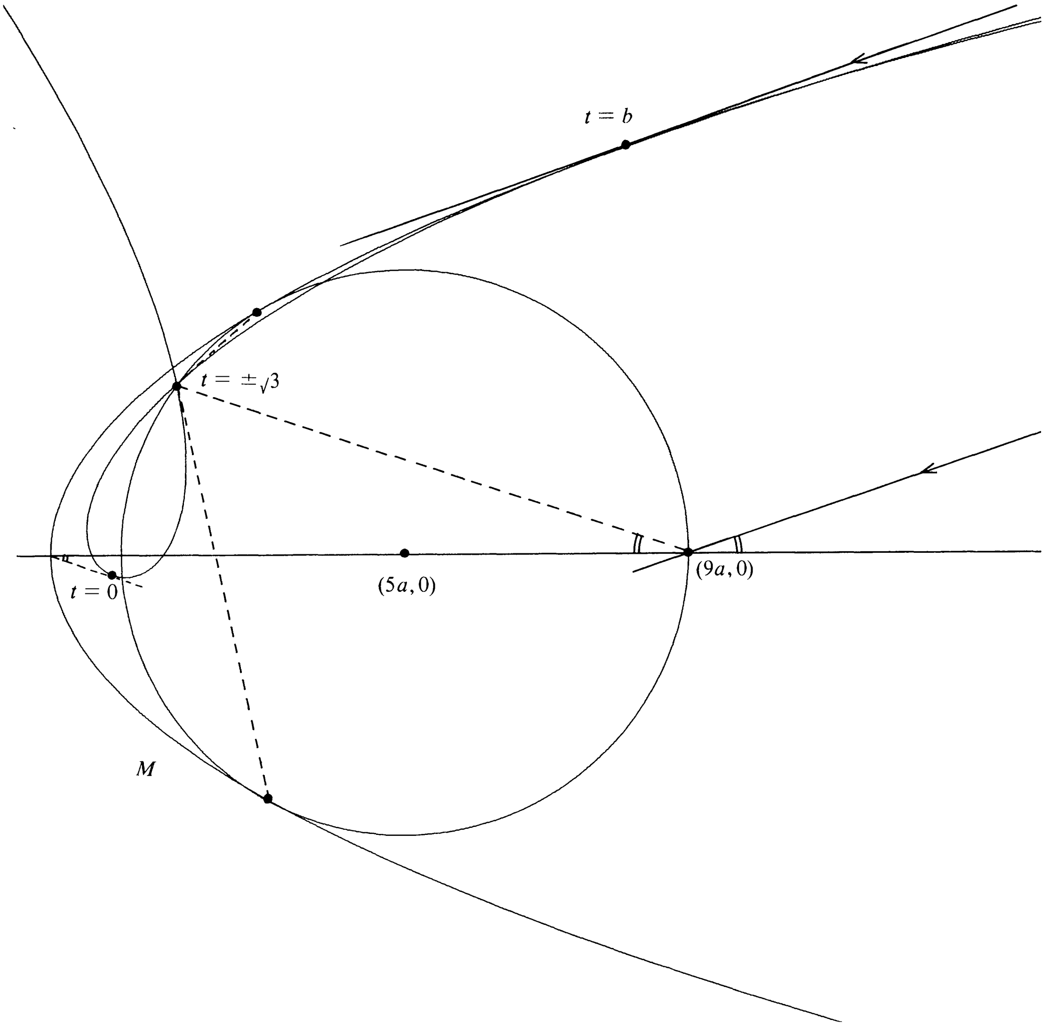


FIG. 18

We now wish to find the relation between the orders of contact of two curves M_1 and M_2 at P and of their orthotomics W_1 and W_2, relative to the same point S, at the corresponding point Q. Clearly, if M_1 and M_2 have the same tangent at P, then, since QP is normal to both W_1 and W_2, the orthotomics will have the same tangent at Q. In fact, in [**1**] we show that the orthotomic is a kind of dual curve, and from this it can be shown that the orders of contact will always coincide. However, we give here a direct and elementary proof for the cases which concern us. First, two preliminary results.

(4.2) PROPOSITION. *Suppose that the curvature κ of M at P is nonzero. Let p be the distance of the point S from the tangent to M at P, and $r \neq 0$ the distance SP. Then the curvature of the orthotomic W at Q is*

$$\kappa_0 = \frac{2r^2\kappa - p}{2r^3\kappa}.$$

Proof. The numbers p, r are "pedal coordinates" for the curve M (see for example [**9**], [**11**]). Taking S as origin (see Fig. 19) and parametrizing M by arc length s (differentiation with respect

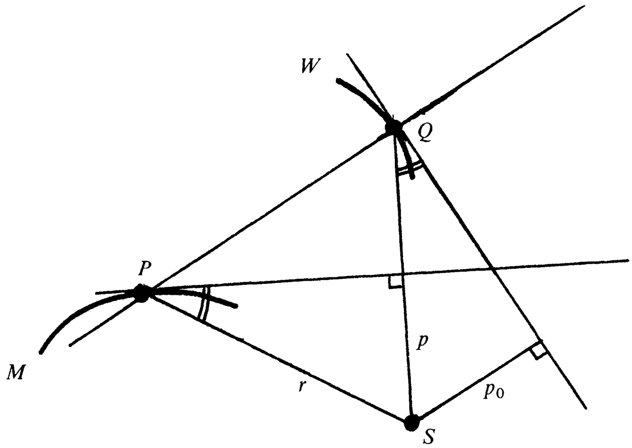


FIG. 19

to s being denoted by a dot), we can use r as a local parameter on M near P provided $\dot{r} \neq 0$ at P, i.e., (with $S \neq P$) provided $x\dot{x} + y\dot{y} \neq 0$ at P. This says that S is not on the normal to M at P. Assuming this, we have

$$\kappa = \frac{1}{r}\frac{dp}{dr} = \frac{1}{r}p',$$

say, on M and close to P. (See [3], [9, p. 24] for the formula for κ.) Now write p_0, r_0 for the pedal coordinates of Q on W (relative to the same point S). Clearly $r_0 = 2p$; also from similar triangles (using the equality of the marked angles) we have $p_0 = 2p^2/r$. We can use r_0 as a local parameter on W near Q provided S is not on the normal QP, but that follows from the assumption above. Thus, writing κ_0 for the curvature of W,

$$\kappa_0 = \frac{1}{r_0}\frac{dp_0}{dr_0} = \frac{1}{r_0}\frac{p_0'}{r_0'}$$

(dashes denoting derivatives with respect to r), and this comes to the formula required. Finally the functions involved are continuous; so the formula must remain valid when S is on the normal at P (not at P itself). (We can avoid an appeal to continuity by doing the calculation in terms of cartesian coordinates x and y, but the calculation is a good deal longer.) □

(4.3) PROPOSITION. *Let s denote arc length on M and let t denote arc length on W. Then, with appropriate orientation,*

$$\frac{dt}{ds} = 2\kappa r,$$

where κ = curvature of M.

Proof. Taking S as origin, with $P = (x, y)$, we have

$$r^2 = x^2 + y^2, \quad \text{so that} \quad r\dot{r} = x\dot{x} + y\dot{y}.$$

Hence

$$\dot{r}^2 = (x\dot{x} + y\dot{y})^2/r^2 = (r^2 - p^2)/r^2.$$

(Recall $p = x\dot{y} - \dot{x}y$ and $\dot{x}^2 + \dot{y}^2 = 1$.)

Similarly,

$$(dr_0/dt)^2 = (r_0^2 - p_0^2)/r_0^2 = (r^2 - p^2)/r^2,$$

where p_0, r_0 are as in the proof of (4.2). Thus

$$\left(\frac{dt}{ds}\right)^2 = \left(\frac{dt}{dr_0}\right)^2\left(\frac{dr_0}{dr}\right)^2\left(\frac{dr}{ds}\right)^2$$

and the middle term is $4\kappa^2 r^2$. The result now follows by orienting W appropriately. □

(4.4) PROPOSITION. *Two curves M_1 and M_2 have at least 3-point contact at P if, and only if, their orthotomics W_1 and W_2 with respect to S have at least 3-point contact at Q. (We assume, as usual, that our curves M_i have nonzero curvature at P.)*

Also, M_1 and M_2 have at least 4-point contact at P, if, and only if, W_1 and W_2 have at least 4-point contact at Q.

Proof. It follows from (4.2) that if $\kappa(M_1) = \kappa(M_2)$ at P then $\kappa(W_1) = \kappa(W_2)$ at Q, and conversely. Differentiating the formula of (4.2) with respect to s and using (4.3), we find, in the notation of (4.3),

$$\frac{d\kappa(W)}{dt} = \frac{-3r^2\dot{r}\kappa^2 + 3p\dot{r}\kappa + pr\dot{\kappa}}{4r^5\kappa^3}.$$

Recall that $\dot{r}^2 = (r^2 - p^2)/r^2$ from the proof of (4.3); this will have the same value for M_1 and M_2 at P when they have the same tangent there. If $\kappa(M_1) = \kappa(M_2)$ at P but the derivatives with respect to arc lengths are different at P, then the formula above shows that the derivatives of $\kappa(W_1)$ and $\kappa(W_2)$ at Q, with respect to arc lengths, will be different. Appealing to the criterion (i) of (4.1) we find that, if M_1 and M_2 have 3-point contact at P, then W_1 and W_2 have 3-point contact at Q. The converse is similar. □

Another interpretation of (4.3) is as follows. Let M be parametrized by arc length s, and write $(X(s), Y(s))$ for the point of W corresponding to the point with parameter s on M. Then (4.3) shows that

$$\dot{X}^2 + \dot{Y}^2 = 4\kappa^2 r^2.$$

It follows from this that $\dot{X}$ and $\dot{Y}$ are never simultaneously zero so long as $\kappa \neq 0$ on M and the source S does not lie on M. Hence:

(4.5) PROPOSITION. *The orthotomic W of M relative to S is an immersed curve in the plane, which can be properly parametrized by arc length on M, provided M has nowhere zero curvature and S does not lie on M.* □

Thus with the assumptions of (4.5) the orthotomic W has no cusps (though it may have double points (nodes)), and its evolute, namely, the caustic of M relative to S, is accordingly a well-defined curve.

The first author gratefully acknowledges the financial support of the Science Research Council.

References

1. J. W. Bruce, P. J. Giblin, and C. G. Gibson, On caustics by reflection, Topology (to appear).
2. A. Cayley, A memoir upon caustics, Philos. Trans. Roy. Soc. London, 147 (1857) 273–312; Collected Works, vol. 2, pp. 336–380.
3. R. H. Fowler, Elementary Differential Geometry of Plane Curves, Cambridge University Press, Tracts in Mathematics and Mathematical Physics No. 20, 1920.
4. E. Goursat, A Course of Mathematical Analysis, Ginn, Cambridge, Mass., 1904.
5. J. Guckenheimer, The cusps of Zeeman's catastrophe machine, Topology, 16 (1977) 177–180.
6. R. S. Heath, Geometrical Optics, Cambridge University Press, 1887.
7. R. A. Herman, Geometrical Optics, Cambridge University Press, 1900.
8. C. Huygens, Abhandlung über das Licht (written 1678, published 1690, German transl., Leipzig, 1913).
9. J. D. Lawrence, A Catalog of Special Plane Curves, Dover, New York, 1972.
10. G. H. Light, The existence of cusps on the evolute at points of maximum and minimum curvature on the base curve, this MONTHLY, 26 (1919) 151–154.

11. E. H. Lockwood, A Book of Curves, Cambridge University Press, 1961.

12. S. H. Schot, Geometrical properties of the penosculating conics of a plane curve, this MONTHLY, 86 (1979) 449–457.

13. D. M. Y. Sommerville, Analytical Conics, G. Bell, London, 1924.

14. C. Taylor, The Elementary Geometry of Conics, with a Chapter on the Line Infinity, D. Bell, Cambridge, 1891.

15. R. C. Yates, Curves and Their Properties, National Council of Teachers of Mathematics, Washington, D.C., 1952.

STRANGE TERRAIN—NONARCHIMEDEAN SPACES

LAWRENCE NARICI

Mathematics Department, St. John's University, Jamaica, NY 11439

EDWARD BECKENSTEIN

Mathematics Department, St. John's University, Staten Island, NY 10301

Quite some time ago I ("I" meaning the first author here) read a science fiction story, the record of which is still probably in the archives of the Brooklyn Public Library's Children's Section, in which travelers went to a strange planet and talked of a color they had never seen. I closed my eyes and tried to picture such a thing. Only greens and browns and purples came, all blending into black eventually, and finally I gave up. My conclusion at the time was that I could only picture colors I had already seen, or very close to them anyway, and I wondered if the same was true of shapes. At that I went through a series of blobs, after passing through the geometric shapes quickly.

The point of this paper is to describe "shapes" in nonarchimedean normed spaces and to discuss briefly what functional analysis is like when the underlying field is not **R** or **C**, but a field F with a nonarchimedean absolute value.

To begin, note some things about the "classical" (real or complex) spaces. One may, for example, casually choose a positive scalar, as when defining convexity; one may casually make use of the fact that **R** is totally ordered. Not only is there the rich topological structure, including local compactness and connectedness, there is also **C**'s algebraic closedness and the whole glorious edifice of holomorphic function theory. Imagine Banach algebra theory without the complex numbers underneath. Can anything remain?

In order to discuss *nonarchimedean* functional analysis, one must start with the underlying field. Fields with absolute values are defined and discussed briefly in Section 1, with special attention given to nonarchimedean valuations, absolute values satisfying $|a+b| \leq \max(|a|,|b|)$ instead of the triangle inequality. For what comes after that, these are the only kinds of absolute values that are considered, the reason being that fields with *archimedean* valuations are just subfields of **C** (see Section 1). Because of this specialization, it is not a more general functional analysis that develops, just a different one. Some of the basic properties of nonarchimedean valued fields—their topological structure and what analysis is like in them—are discussed in Sections 1 and 2. By Section 3 we have worked our way up far enough from the ground field to speak of Banach spaces. We mention that many of the central results of classical Banach space

---

Lawrence Narici and Edward Beckenstein each studied mathematics at the Polytechnic Institute of New York (formerly, "of Brooklyn") with George Bachman. Their previous collaborations in functional analysis include the books *Functional Analysis and Valuation Theory*, with G. Bachman (Marcel Dekker, New York, 1971), and *Topological Algebras*, with C. Suffel (North-Holland, Amsterdam, 1977). Narici is also coauthor (with G. Bachman) of *Functional Analysis* (Academic Press, New York, 1966).—*Editors*

theory (Closed Graph, Open Mapping, and Banach-Steinhaus Theorems) carry over, albeit with very different proofs at times, while the Hahn-Banach theorem fails. The cause of this failure is not in the vector space itself, but rather in the ground field. For *spherically complete* ground fields, all linear spaces over such fields possess a Hahn-Banach theorem. Defining a topological vector space (TVS) over a field with a nonarchimedean valuation presents no problem at all. But the theory of real or complex TVS's is really the theory of locally convex (Hausdorff?) TVS's, and it is in defining convexity that difficulties arise. (In our opinion, without local convexity one may as well deal with topological groups.) Nonarchimedean valued fields are not ordered *a priori* and, moreover, since they are totally disconnected, they *cannot* be totally ordered. Various substitutes for local convexity have been proposed, and one called *F-convexity* is discussed here. Duplication of a Hilbert space theory is rather a lost cause.

The subject of Section 4 is (nonarchimedean) Banach algebras. No field containing **C** can be a normed algebra over **C**; while for nonarchimedean valued fields F, the valuation on F can be extended to any field containing F. This vital difference makes the complex and nonarchimedean theories but distant cousins. In Section 5 we discuss spaces $C(T, F)$ of continuous functions, first just as algebras and then as locally F-convex TVS's. As an algebra some results are needed in the style of Gillman and Jerison before we can get on to discuss analogues of the Nachbin-Shirota theorems concerning barreledness and bornologicity of spaces of continuous functions with compact-open topology. Along the way we also mention measure and integration theory in this context as it applies to getting representation theorems for continuous linear functionals. A little more discussion of measure and integral is given in Section 6. In the sequel, please note, we say a little about some main areas of nonarchimedean functional analysis—we don't say a little about everything—and refer to the books in the bibliography for extensive lists of references.

1. Valued Fields. It was Kürschak who, in 1913, gave the general definition of absolute value. Calling it a *valuation*, he said it is a map $|\ |$ from a field F into the nonnegative reals such that only $|0|$ is 0, $|ab|=|a||b|$, and

$$\text{(T)} \qquad |a+b|\leqslant|a|+|b| \quad \text{for all } a, b \in F.$$

When a field carries an absolute value, it is called a *valued field*. Examples of valuations are provided by the usual absolute values of **R** and **C**. An example of one which is unlike these is given in Example 1 below.

The inequality (T), the triangle inequality, basically makes the shortest distance between two points a straight line and, generally, allows for confirmation of things one would expect to be true from experience with plane geometry or just observation of the world around you. If we take "shape" to mean something like "the totality of relative distances of things from each other (in a given set)," then analogues in valued fields of entities such as circles have many of the properties one expects them to have. But there is another kind of valuation, whose importance Kürschak realized, which satisfies a stronger condition than the triangle inequality, and causes our intuitive expectations to be violated regularly and dramatically (see, e.g., Example 3 below). If (T) is replaced by

$$\text{(ST)} \qquad |a+b|\leqslant \max(|a|,|b|) \quad \text{for all } a, b \in F,$$

the map $|\ |$ is then called a *nonarchimedean* or *ultrametric* valuation. Part of the importance of this type is revealed by a theorem of Ostrowski which shows that fields with nonultrametric valuations are just subfields of **C** with ordinary absolute value ([**1**, p. 27]).

An example of a nonarchimedean valuation is given by one called the *trivial* valuation, the map $|\ |$ on any field F taking everything but 0 into 1; $|0|=0$. It has its uses in counterexamples and makes for provisos such as "where F is nontrivially valued" in many a theorem, but it has other uses as well [**43**], [**56**]. The trivial valuation turns any field into a discrete topological space.

A less trivial example of an ultrametric valuation, and undoubtedly the one that motivated Kürschak to the general consideration, was invented by Hensel in 1907. It is the p-adic valuation.

EXAMPLE 1. *p-Adic Valuations on Q*. If one selects a certain prime p, then any rational number x can be written as $p^m(a/b)$ for some integer m, where a and b are relatively prime to each other and to p. The *p-adic valuation* of x, $|x|_p$, is then defined to be p^{-m}. That it satisfies the strong triangle inequality (ST) is verified in Bachman's book [**1**], which is a pleasant introduction to valuation theory, and in many other books dealing with field theory, commutative algebra, or algebraic number theory. Another good introductory treatment is [**3**].

EXAMPLE 2. *The Field* Q_p. When Q carries a p-adic valuation, it becomes a metric space. As such, it may be completed by a standard process. The complete space which is obtained by this process is easily made into a field Q_p which contains the rationals as a subfield; Q_p is called the (field of) *p-adic numbers*. There is already a certain analogy perceivable between Q_p and $\mathbf{R}$ in that each is realized as a metric completion of the rationals. The analogy goes a little further in that each x in Q_p admits a representation of the decimal type: $x = \sum_{j \geqslant n} a_j p^j$ where n and the a_j are integers [**1**, p. 35]. It is also true even that a p-adic number is a rational number if and only if this decimal representation is periodic. But, though there are similarities, there are significant differences. In Q_5, for example, the equation $x^2 + 1 = 0$ may be solved.

For p-adic number fields there is an extensive literature devoted to differentiability, solving equations, power series, and analytic functions. The "exponential" and "logarithmic" series

$$E(x) = \sum_{n \geqslant 0} (x^n/n!) \quad \text{and} \quad L(1+x) = \sum_{n \geqslant 1} (-1)^{n-1} x^n/n,$$

for example, converge for all $x \in Q_p$, such that $|x|_p < 1$, and have properties which are very similar to their classical counterparts: $E(x+y) = E(x)E(y)$, $L((1+x)(1+y)) = L(1+x) + L(1+y)$, $L(E(x)) = x$, and $E(L(1+x)) = 1+x$. These things, in turn, have many applications. Using the log function above, for example, one can demonstrate the number-theoretic fact [**2**, p. 289] that $2 + 2^2/2 + \cdots + 2^n/n$ is divisible by arbitrarily large powers of 2 for n sufficiently large.

EXAMPLE 3. *Circles and Triangles*. Let F be a field with a nonarchimedean valuation. A *closed circle* (or *disc*) in F is set of the form $C(a, r) = \{x \in F\colon |x - a| \leqslant r\}$ where $a \in F$ and r is a positive number. The *open disc* is defined by using the strict inequality. It is easy to show that if $b \in C(a, r)$ then $C(b, r) = C(a, r)$. In other words, any point in a circle is a center. Similarly, taking "triangle" to mean any triple (a, b, c) of points from F, all triangles are isosceles.

The closed (or open) discs mentioned in Example 3 determine a topology for a nonarchimedean valued field in the usual way. Since each of the discs $C(a, r)$ is an open, as well as a closed, set, ultrametric topologies are seen to be highly disconnected (totally disconnected, to be exact) and 0-dimensional in the sense that there is a base for the topology consisting of sets which are simultaneously closed and open. We abbreviate "closed and open" to just "clopen." A corresponding term used by some French authors is "oufermé" from ouvert and fermé. Some suggestions for Portuguese, German, and Spanish equivalents (solicited at a cocktail party after a conference) were, respectively, "feberto" from fechado and aberto, "geschloffen" from abgeschlossen and offen, and "aberrado" from abierto and cerrado and its resemblance to "aberrated." In fairness, it should be mentioned that some German-speaking mathematicians noticeably recoiled from "geschloffen."

One might be convinced that nonarchimedean valued fields have their uses, but that nonarchimedean distances, nonarchimedean shapes, exist only in the minds of mathematicians. This may not be so. Everett and Ulam [**31**] have conjectured that distances between particles "in the small" (less than 10^{-13} cm) may behave in a nonarchimedean manner. They suggest that it is not practical to assume that there is a basic quantum of distance in the small but that there may be a discrete (infinite) set of possible distances, with arbitrarily small ones possible.

NOTATION. For the rest of the paper, F denotes a field with a nonarchimedean valuation.

2. Analysis in Valued Fields. One can develop an analogue of analytic function theory in nonarchimedean valued fields F [**60**]–[**71**]. A good introduction to what it is like in Q_p is presented in [**1**]. If x and a_n $(n \in \mathbf{N})$ are from Q_p, for example, the power series $f(x) = \Sigma_n a_n x^n$ converges for $|x|_p < r = 1/\limsup_n(|a_n|_p)^{1/n}$. For $f'(x) = \Sigma_n n a_n x^{n-1}$ it is even true that

$$(*) \qquad f'(x) = \lim_{y \to 0} \left(f(x+y) - f(x) \right)/y.$$

More generally in regard to derivatives (cf. [**10**], [**59**]), if U is any subset of a nonarchimedean valued field F without isolated points and $f \colon U \to F$ we say that f is *differentiable at* $x \in U$ if the limit $(*)$ exists. With this definition the characteristic function (F-valued) k_U of any clopen set is differentiable everywhere; in fact, $k'_U = 0$ everywhere. Thus there are nonconstant functions whose derivatives are 0 everywhere. There are even 1-1 functions whose derivatives are 0 everywhere: If the characteristic of F is 2 and $f \colon F \to F$ is the map sending x into x^2, then f is 1-1 $(x^2 = y^2 \Rightarrow x = y)$ while

$$f'(a) = \lim_{x \to a} \frac{x^2 - a^2}{x - a} = \lim_{x \to a} (x - a) = 0$$

at any point a. It is even easy to get functions which are not differentiable anywhere even though they are continuous everywhere. For example, letting U denote the closed unit disc in Q_p, for any p, the map $f \colon U \to Q_p$ sending $\Sigma_n a_n p^n \to \Sigma_n a_n p^{2n}$ $(a_n \in \{0, \ldots, p-1\})$ is continuous everywhere but differentiable nowhere [**26**].

In Archimedean (= ordinary) analysis, functions which are derivatives—i.e., functions which have antiderivatives—are known not to have jump discontinuities and to be pointwise limits of continuous functions. But not even both these conditions are enough to guarantee that the functions have an antiderivative: sufficient conditions are simply not known. In the nonarchimedean case the situation is much simpler: letting U be a subset of F without isolated points, $f \colon U \to F$ has an antiderivative if and only if f is the pointwise limit of continuous functions [**5**, p. 283].

3. Banach Spaces. The definition of a nonarchimedean Banach space X over a nonarchimedean valued field F is formally the same as for the real case except that the valuation and norm satisfy the strong triangle inequality: $\|x+y\| \leqslant \max(\|x\|, \|y\|)$. Since we will only deal with nonarchimedean Banach spaces, we drop the "nonarchimedean" in what follows and say just "Banach space." We also require that the valuation on F be nontrivial from this point on. The reason for this is to be able to say that a linear map between Banach spaces is continuous if and only if it is bounded, a result which may fail if the valuation on f is trivial. Even excluding trivial valuations on F, there may not be any unit vectors in X. A vector x cannot be divided by $\|x\|$, because $\|x\|$ is a real number, not a scalar. Because of this deficiency, even one-dimensional Banach spaces need not be isometrically isomorphic, although they must be linearly homeomorphic. A question that is still open is: Given a Banach space X over F, is there an equivalent norm $\| \ \|'$ for X such that $\|X\|' = |F|$?

Consider the following formulas for the norm of a bounded linear map $A : X \to Y$ where X and Y are Banach spaces.

$$\|A\| = \sup\{\|Ax\|/\|x\| : 0 < \|x\| \leqslant 1\} \tag{1}$$

$$\|A\| = \sup\{\|Ax\| : \|x\| = 1\} \tag{2}$$

$$\|A\| = \sup\{\|Ax\| : 0 \leqslant \|x\| \leqslant 1\} \tag{3}$$

If F does not carry the trivial valuation, (1) holds, but (2) and (3) can still fail [**4**, p. 75].

Many results about classical and nonarchimedean Banach spaces look the same—have exactly the same formal statement—but have entirely different proofs. The theorem that locally compact

Banach spaces must be finite-dimensional is a good case in point. The classical argument involves the casual use of unit vectors; the nonarchimedean proof does not (it can't) and proceeds by considering various cases and vectors whose norm is close to one. *Then* it follows that locally compact Banach spaces X must be finite-dimensional and the underlying field F must be locally compact [**4**, p. 70 ff.].

Some seminal results of classical Banach space theory which remain true for nonarchimedean Banach spaces are the Closed Graph, Banach-Steinhaus, and Open Mapping theorems. In some ways the situation is more interesting when a theorem does *not* carry over, as happens with the Hahn-Banach theorem. In nonarchimedean Banach spaces it may indeed be impossible to extend a given continuous linear functional from a subspace to the whole space. The fault, though, lies not with the vector space X but with the underlying field F. After some success in special cases—e.g., Cohen's [**25**] in 1948, when the underlying field was complete and had a discrete valuation—attention was finally focused on the underlying field F, and Ingleton [**35**] (with some inspiration from Nachbin [**38**]) proved the central result: [**4**, p. 78].

THE NONARCHIMEDEAN HAHN-BANACH THEOREM. *Let X be a normed space and Y a nonarchimedean normed space over F (F may even be trivially valued). A continuous linear map A defined on a subspace M of X into Y may be extended to a continuous linear map $A^* : X \to Y$ with the same norm (as in Equation* (1)) *iff Y is spherically complete in the sense defined below.*

Spherical completeness (of a metric space) is a stronger notion than completeness. Taking completeness to mean that every nested sequence of closed balls whose diameters shrink to 0 have nonempty intersection, spherical completeness makes the same demand but drops the requirement that the diameters go to 0. Naturally, any spherical complete field is complete, but the converse is false. In any case Ingleton's version covered the known special cases and completely solved the problem.

There are other ways to describe spherical completeness (four equivalents are discussed in Chapter 2 of [**4**]), including one which involves something like Cauchy sequences. A notion that should be dispelled right away is the idea that spherical completeness should be substituted in nonarchimedean theorems where completeness appears in the Archimedean ones. It plays practically no role in nonarchimedean Banach algebras and none whatsoever in nonarchimedean measure theory. In some ways, in fact, it hurts, viz., if F is spherically complete, then no infinite-dimensional Banach space over F is reflexive [**5**, Chapter 4]. If F is *not* spherically complete, then c_0 and l^∞ (defined as exact analogues of their classical counterparts) *are* reflexive.

In addition to its requirements about complex conjugates, the inner product $(,)$ on a classical Hilbert space must satisfy the condition $(x, x) \geqslant 0$ for every x. Assuming that a potential inner product on a linear space X over a nonarchimedean valued field F is scalar-valued, the "$\geqslant 0$" part of the definition will not transfer. Moreover, since nonarchimedean fields are totally disconnected, they *cannot* be totally ordered. The consequence of this is that there is really no such thing as a nonarchimedean Hilbert space (but, see [**36**]), though there is a notion of "orthogonality" in nonarchimedean Banach spaces. Using this notion of orthogonality, there is such a thing as an "orthogonal base" which has many of the properties that its namesake in ordinary Hilbert spaces does, though perhaps the analogy with Schauder bases is more appropriate [**5**, Chapter 5].

The lack of ordering on F also makes itself felt when one tries to find an analogue for convexity. A notion that is used is called "F-convexity" and is defined as follows: A collection S of vectors is *absolutely F-convex* if $ax + by \in S$ for $|a|, |b| \leqslant 1$ and $x, y \in S$; translates $w + S$ of such sets are then called *F-convex*. Defining "topological vector space" in formally the same way as for real or complex spaces, a topological vector space would be "locally F-convex" if its topology had a base of F-convex sets at 0. A seminorm p is *nonarchimedean* if it is a seminorm in the usual sense but also satisfies $p(x + y) \leqslant \max(p(x), p(y))$. The connection between F-convex sets and nonarchimedean seminorms is quite similar, although a little more complicated, to the Archimedean case. Of course once a notion of convexity has been introduced, the way to defining

such things as F-barrels is clear. Using these notions in spaces of continuous functions, even analogues of the Nachbin-Shirota theorems concerning the barreledness and bornologicity of such spaces may be obtained, as is discussed in Section 5.

4. Banach Algebras. In classical Banach spaces, real and complex spaces are treated on pretty much equal footing. In Banach algebras, however, the real space is decidedly weaker than its complex counterpart. One reason for this is the interest in the spectra of elements, which, as Rickart [**44**, p. 27] observes, "is perhaps the most important notion in the theory of Banach algebras." A seminal result for complex Banach algebras is that no element can have an empty spectrum and one way to prove it [**17**, p. 212] draws on the complex substructure by using Liouville's theorem about the constancy of bounded entire functions. This is not the only way to do it, as Warner's recent proof [**57**] shows; his proof does not rely on holomorphic function theory at all. Even so, the underlying field still has to be **C** or the theorem fails.

A consequence of the nonemptiness of spectra is that no field can contain **C**, which is a normed algebra over **C**. This, in turn, gives rise to a way of topologizing the maximal ideals of a complex Banach algebra—known as the Gelfand theory—with many interesting consequences. For nonarchimedean Banach algebras (same definition, formally, but for the strong triangle inequality) the situation is reversed: there are loads of fields containing F which are normed algebras over F. By a theorem due to Krull, a nonarchimedean valuation on a field can *always* be extended to a field containing it. This stark difference is the nail by which the kingdom is lost. Maybe "lost" is too strong a word. The theory that develops is very different from its classical counterpart; in some cases, searching for nonarchimedean analogues of results for complex algebras leads to barking up the wrong tree—one might be better off not knowing the result for complex algebras. There are analogues of the B^*-algebra (the "V^*-algebra" [**4**, § 4.8] or the C^*-algebra [**5**, p. 213]; see also [**37**] and [**54**]), some semblance of a Gelfand theory [**4**, Chapter 4], [**46**], and development of individual Banach algebras such as group algebras $L(G)$ [**5**] and Banach algebras of continuous F-valued functions. There is a Stone-Weierstrass theorem due to Kaplansky [**4**, p. 159] for continuous functions whose range is in a nonarchimedean valued field. Indeed, there is an approximation theory for functions with values in a nonarchimedean valued field (see [**23**], [**24**], [**41**] and [**42**]).

Thus far, perhaps, those who have investigated nonarchimedean Banach algebras have been too heavily influenced by the known results for complex Banach algebras.

5. Spaces of Continuous Functions. $C(T, F)$ denotes the F-algebra (pointwise operations) of continuous functions mapping the topological space T into the nonarchimedean valued field F. Before endowing $C(T, F)$ with a topology, let us note some facts about it. (1) To get T's topology generated by $C(T, F)$, it is necessary and sufficient that T be a 0-dimensional Hausdorff space; that is to say, T must be ultraregular. The reason for the desirability of this condition is to make for enough continuous functions to guarantee sufficient interaction between T and the continuous F-valued functions on T. If T, say, was indiscrete, $C(T, F)$ would consist only of constants and it would be impossible to learn anything else about T's fine structure by just looking at the continuous functions on T. (2) For the algebra $C(T, \mathbf{R})$ of real-valued continuous functions, where T is a completely regular Hausdorff space, the maximal ideals of $C(T, \mathbf{R})$ are in 1-1 correspondence with the points of βT, the Stone-Čech compactification of T. Not all of these maximal ideals determine real-valued homomorphisms of $C(T, \mathbf{R})$; the collection vT of points of βT which do is called the *repletion* (*real-compactification*) of T. βT and vT are each "completions" of T with respect to the initial uniformities determined respectively by $C_b(T, \mathbf{R})$—the *bounded* continuous functions—and $C(T, \mathbf{R})$ (see [**14**], [**17**, Chapter 1], and [**15**]). How much of this remains when **R** is replaced by F and T by an ultraregular space? The answer is: practically everything [**15**]. Instead of the Stone-Čech compactification of T, we consider the completion $\beta_F T$ with respect to the initial uniformity determined by $C(T, F)$; similarly we use $v_F T$, the completion of T with respect

to the initial uniformity determined by $C(T, F)$ as the substitute for vT. In fact (see [**15**]) we can consider βT and vT as special cases of completions $\beta_F T$ and $v_F T$ with $F = \mathbf{R}$. Having done this, the maximal ideals of $C(T, F)$ are determined by the points of $\beta_F T$, the F-valued homomorphisms of $C(T, F)$ by $v_F T$.

Now let us topologize the spaces of continuous functions and return to functional analysis proper. The way to do this is the same whether the functions are real-valued or F-valued. The most popular topology for this purpose is the compact-open topology (also known as the topology of compact convergence and the topology of uniform convergence on compact sets), that topology generated by the seminorms p_K:

$$p_K\colon C(T, F) \to \mathbf{R}^+ \text{ (the nonnegative reals)}$$

$$x \to \sup |x(K)|$$

for each compact subset K of T. $C(T, F)$ so topologized is denoted by $C(T, F, c)$. The seminorms p_K are nonarchimedean in the sense used in Section 3 and $C(T, F, c)$ is a locally F-convex space. As the p_K's are also submultiplicative, $C(T, F, c)$ is a topological algebra as well.

Two of the most prominent results about $C(T, \mathbf{R}, c)$, first proved by Nachbin [**39**] and Shirota [**48**], characterize those T's for which (A) $C(T, \mathbf{R}, c)$ is barreled and (B) $C(T, \mathbf{R}, c)$ is bornological. Letting T be a completely regular Hausdorff space, the exact statements are (cf. [**17**], Chapter 2):

(A) $C(T, \mathbf{R}, c)$ is barreled if and only if for each closed noncompact subset S of T there is some $x \in C(T, \mathbf{R})$ which is unbounded on S.

(B) $C(T, \mathbf{R}, c)$ is bornological if and only if T is replete, i.e., $T = vT$.

The analogous results when $\mathbf{R}$ is replaced by a nonarchimedean valued field F are [**14**], [**15**]:

(A′) If F is spherically complete then $C(T, F, c)$ is F-barreled if and only if on each closed noncompact subset S of T there is some $x \in C(T, F)$ which is unbounded on S.

(B′) If F has nonmeasurable cardinal* then $C(T, F, c)$ is F-bornological if and only if T is F-replete, i.e., $T = v_F T$.

The proofs of these results are similar in appearance to the classical arguments. A major difference is that a good deal of machinery for the classical arguments already exists and only needs to be used, things like measure and integration theory, for example. To get (A′), an appropriate measure and integral must be invented and developed. A *bounded measure* is taken to be a map μ mapping the clopen subsets of T into the field F which is finitely additive and bounded in absolute value. After introducing a suitable integral involving such μ's, a representation theorem for continuous linear functionals on $C(T, F, c)$ in terms of integrals is obtained—when T is compact, there is an isometric isomorphism between the continuous dual $C(T, F, c)'$ of $C(T, F, c)$ and the normed linear space $M(T)$ of bounded measures on T; when T is not compact, the correspondence is between $C(T, F, c)'$ and a subspace of $M(T)$ ([**13**] and [**5**, p. 270, and Notes, Ch. 7]). A few more things still have to be done, but these are the main ingredients used to establish (A′).

The well-beaten paths through the theory of rings of continuous functions à la Gillman and Jerison serve as a necessary backdrop for (B) and parts of an analogous theory for rings of continuous functions with values in a topological field need to be developed to get (B′). The details are in [**15**] with further developments in [**32**]–[**34**], characterizing, among other things, ultrabornologicity and semi-bornologicity of $C(T, F, c)$.

6. Measure and Integral. To get a nonarchimedean measure theory we use set functions μ that

*A measure μ defined on the class $P(X)$ of all subsets of a set X is concentrated at $x \in F$ if $\mu(A) = 1$ if $x \in A$ and $\mu(A) = 0$ otherwise. Such a μ is called a *point mass*. A set X is of *measurable cardinal* if there exists a countably additive measure μ defined on $P(X)$ whose range is $\{0, 1\}$ which is not a point mass. Such sets must be so enormous that their very existence is in doubt. For some further discussion, see [**17**, pp. 33, 34, 43].

are F-valued, rather than real-valued. The domain for the nonarchimedean measures defined in the preceding section has been the collection C of clopen subsets of an ultraregular space T, a connotation which T carries throughout this section. In Section 5 we mentioned how a measure of this type is used to create a representation theory for spaces of continuous functions and, ultimately, to characterize those spaces T for which $C(T, F, c)$ is barreled. If we consider bounded finitely additive set functions $\mu: C \to F$ which may only assume the values 0 and 1 and which are *monotone* in the sense that (clopen) subsets of sets of measure 0 must also be of measure 0, we denote this collection by $M_0(T)$. (If the characteristic of F is not 2, then each bounded finitely additive set function μ defined on C whose range is $\{0, 1\}$ is monotone.) $M(T)$ denotes the collection of all bounded finitely additive F-valued set functions defined on C, as in Section 5. Suitably topologized, $M_0(T) = \beta_F T$ and $v_F T$ consists of the countably additive members of $M_0(T)$ **[14]**, **[15]**.

Something different to be noted at this point is that the domains of the so-called measures mentioned so far were not σ-algebras but the ring C of clopen subsets of (the ultraregular space) T. A reason for this is that F-valued set functions on σ-algebras are sort of trivial (see below). There is a basic incompatibility between σ-algebras and nonarchimedean valuations **[5**, p. 116]: a countably additive F-valued set function $\mu: H \to F$, where H is a σ-algebra of subsets of a set, is purely atomic in the sense that it must essentially be concentrated in a denumerable collection of sets A_n (called *atoms*). The difficulty does not lie with the countable additivity requirement on μ, for on rings of sets that aren't σ-algebras there are nontrivial countably additive F-valued measures. In any case, though it is quite different from the classical theory, there is quite a bit of nonarchimedean measure and integration theory, mostly for the case when the ring of sets is C. Of particular interest are Van der Put's "tight" measures (see **[52]** and their development in Van Rooij's book **[5]**). A convolution may be defined for them which makes a Fourier theory possible, the tight measures playing the part that bounded measures play in the Archimedean theory. There is also a nonarchimedean harmonic analysis **[5]**, **[9]**, **[58]**.

The references below, together with the bibliographies of **[4]**, **[5]**, and **[8]**, provide a fairly decent list of what's been done in the areas discussed in this paper (i.e., it is quite incomplete for items not discussed at length here, such as analytic function theory) up to about 1972.

References

Books

1. G. Bachman, Introduction to p-adic Numbers and Valuation Theory, Academic Press, New York, 1964.

2. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.

3. K. Mahler, Introduction to p-adic Numbers and Their Functions, Cambridge University Press, Cambridge, 1973.

4. L. Narici, E. Beckenstein, and G. Bachman, Functional Analysis and Valuation Theory, Marcel Dekker, New York, 1971.

5. A. Van Rooij, Nonarchimedean Functional Analysis, Marcel Dekker, New York, 1978.

Reports and Theses

6. W. Blum, Über nichtarchimedische Banachalgebren, dissertation, Karlsruhe, 1970.

7. G. Isac and G. Marinescu, Analiza pe corpuri ultrametrice, Editura Academiei Republicii Socialiste Romania, Bucharest, 1976.

8. A. F. Monna, Analyse Non-archimédienne, Springer-Verlag, Berlin-Heidelberg-New York, 1970.

9. W. H. Schikhof, Nonarchimedean harmonic analysis, dissertation, Nijmegen, 1967.

10. ______, Nonarchimedean calculus, Report 7812, August, 1978, Math. Inst. Katholieke Universiteit, Toernooiveld, 6525 ED Nijmegen.

11. A. Van Rooij, Notes on p-adic Banach spaces, Reports 7633 and 7725, Math. Inst., Katholieke Universiteit, Nijmegen, 1976, 1977.

12. ______, Nonarchimedean representations of compact groups, Report 7704, Math. Inst., Katholieke Universiteit, Nijmegen, 1977.

Papers and Other Literature

13. G. Bachman, E. Beckenstein and L. Narici, Function algebras over valued fields and measures, I-IV, Atti Accad. Naz. Lincei Rend., Cl. Sci. Fis. Mat. Natur., 51 (1971) 293–300; 52 (1972) 120–125; 52 (1972) 349–358; 52 (1972) 840–845.

14. ______, Some recent developments on repletions and Stone-Čech compactifications of 0-dimensional spaces, in General Topology and Its Applications, Second Pittsburgh International Conference, R. Alo, ed., Springer-Verlag, New York, 1974, pp. 310–321.

15. G. Bachman, E. Beckenstein, L. Narici, and S. Warner, Rings of continuous functions with values in a topological field, Trans. Amer. Math. Soc., 204 (1975) 91–112.

16. E. Beckenstein, On regular nonarchimedean Banach algebras, Arch. Math., 19 (1968) 423–427.

17. E. Beckenstein, L. Narici, and C. Suffel, Topological Algebras, North-Holland, Amsterdam, 1977.

18. E. Beckenstein, G. Bachman and L. Narici, Topological algebras of continuous functions over valued fields, Studia Math., 48 (1973) 119–127.

19. ______, Function algebras over valued fields, Pacific J. Math., 44 (1973) 45–58.

20. W. Blum, Über kommutative nichtarchimedische Banachalgebren, Arch. Math., 24 (1973) 493–498.

21. ______, Über elementare Funktionen in nichtarchimedischen Banachalgebren, Manuscripta Math., 12 (1974) 329–338.

22. N. Bourbaki, Espaces vectoriels topologiques, Hermann, Paris, 1953.

23. J. Carneiro, Nonarchimedean weighted approximation, in Approximation Theory and Functional Analysis, J. Prolla, ed., North-Holland, Amsterdam, 1979, pp. 121–131.

24. ______, Aproximacão ponderada não-arquimediana, An. Acad. Brasil. Cienc., 50 (1978) 1–34.

25. I. S. Cohen, On nonarchimedean normed spaces, Indag. Math., 10 (1948) 244–249.

26. J. Dieudonné, Sur les fonctions continues p-adiques, Bull. Sci. Math., 68 (1944) 79–95.

27. A. Escassut, Spectres algébriques et idempotents associés dans une algèbre de Banach ultramétrique, C. R. Acad. Sci. Paris Sér. A, 275 (1972) 1285–1287.

28. ______, Algèbres de Banach ultramétriques et algèbres de Krasner-Tate, Astérisque, 10 (1973) 1–107.

29. ______, Algèbres de Banach ultramétriques, Bull. Soc. Math. France Mém., 39–40 (1974) 39–54.

30. ______, Propriétés spectrales en analyse nonarchimédienne, C. R. Acad. Sci. Paris Sér. A, 278 (1974) 1387–1389.

31. C. J. Everett and S. Ulam, On some possibilities of generalizing the Lorentz group in the special relativity theory, J. Comb. Theory, 1 (1966) 248–270.

32. W. Govaerts, Bornological spaces of nonarchimedean valued functions, C. R. Math. Rep. Acad. Sci. Canada, 1 (1979) 233–235.

33. ______, Bornological spaces of nonarchimedean valued functions with the point-open topology, Proc. Amer. Math. Soc, 72 (1978) 571–575.

34. ______, Bornological spaces of nonarchimedean valued functions with the compact-open topology, Proc. Amer. Math. Soc., 78 (1980) 132–134.

35. A. Ingleton, The Hahn-Banach theorem for nonarchimedean fields, Proc. Cambridge Philos. Soc., 48 (1952) 41–45.

36. G. Kalisch, On p-adic Hilbert spaces, Ann. of Math., 48 (1947) 180–192.

37. G. Murphy, Commutative nonarchimedean C^*-algebras, Pacific J. Math., 78 (1978) 433–446.

38. L. Nachbin, A theorem of the Hahn-Banach type for linear transformations, Trans. Amer. Math. Soc., 68 (1950) 28–40.

39. ______, Topological vector spaces of continuous functions, Proc. Nat. Acad. Sci. U.S.A., 40 (1954) 471–474.

40. L. Narici, On nonarchimedean Banach algebras, Arch. Math. 19 (1968) 428–435.

41. J. Prolla, Approximation of Vector Valued Functions, North-Holland, Amsterdam, 1977.

42. ______, Nonarchimedean function spaces, in Linear Spaces and Approximation Theory, ed. by P. Butzer and B. Sz.-Nagy, ISNM vol. 40, Birkhäuser Verlag, Basel-Stuttgart, 1978. pp. 101–117.

43. P. Robert, On some nonarchimedean normed linear spaces, Comp. Math., 19 (1968) 1–77.

44. C. Rickart, General Theory of Banach Algebras, Van Nostrand, Princeton, N.J., 1960.

45. H. Schaefer, Topological Vector Spaces, Springer-Verlag, New York, 1971.

46. N. Shilkret, Nonarchimedean Gelfand theory, Pacific J. Math., 32 (1970) 541–550.

47. ______, Nonarchimedean Banach algebras, Duke Math. J., 37 (1970) 315–322.

48. T. Shirota, A class of topological spaces, Osaka Math. J., 4 (1952) 23–40.

49. K. Tsurumi, On Baire's functions in nonarchimedean valued fields, Indag. Math., 37 (1975) 273–283.

50. ______, The characterization of Baire functions of class 1 over a nonarchimedean valued field, Arch. Math., 33 (1979) 165–169.

51. ______, Pointwise compactness in function spaces over nonarchimedean valued field (to appear).

52. M. Van der Put, Reflexive nonarchimedean Banach spaces, Indag. Math., 32 (1970) 341–352.
53. A. Van Rooij and W. Schikhof, Nonarchimedean integration theory, Indag. Math., 31 (1969) 190–199.
54. ______, Nonarchimedean commutative C^*-algebras, Indag. Math., 35 (1973) 381–389.
55. J. Van Tiel, Espaces localement K-convexes, I-III, Indag. Math., 27 (1965) 249–258; 259–272; 273–289.
56. S. Warner, Locally compact vector spaces and algebras over discrete fields, Trans. Amer. Math. Soc., 130 (1968) 463–493.
57. ______, A new approach to Gelfand-Mazur theory and the extension theorem, Mich. Math. J., 26 (1979) 13–17.
58. C. Weisman, Nonarchimedean harmonic analysis and analytic functions, Math. Ann., 210 (1974) 227–231.
59. ______, On p-adic differentiability, J. Number Theory, 9 (1977) 79–86.

Nonarchimedean Analytic Function Theory

60. W. W. Adams and E. G. Strauss, Nonarchimedean analytic functions taking the same values at the same points, Ill. J. Math., 15 (1971) 418–424.
61. Y. Amice, Interpolation p-adique, Séminaire Dubreil-Pisot, 1963/64.
62. D. N. Clark, A note on the p-adic convergence of solutions of linear differential equations, Proc. Amer. Math. Soc., 17 (1966) 262–269.
63. B. Dwork, On the zeta function of a hypersurface, Inst. Hautes Etudes Sci. Publ. Math., Paris, no. 12 (1962) 5–68.
64. ______, On p-adic differential equations I-IV, Bull. Soc. Math. France Mém., 39–40 (1974) 27–37; Ann. of Math., 98 (1973) 366–376; Invent. Math., 20 (1973) 35–45; Ann. Sci. Ecole Norm. Sup., (4) 6 (1973) 295–316.
65. J. de Groot, Bemerkung über die analytische Forsetzung in bewerteten Körpern, Indag. Math., 4 (1942) 120–122.
66. H. Grauert and R. Remmert, Nichtarchimedische Funktionentheorie, Arbeitsgemeinschaft für Forschung Nordheim-Westfallen, Wiss. Abhandl., 33 (1966) 393–476.
67. M. Lazard, Les zéros d'une fonction analytique d'une variable sur un corps valué complet, Inst. Hautes Etudes Sci. Publ. Math., Paris, no. 14 (1962).
68. T. T. Raghunathan, On the space of entire functions over certain nonarchimedean fields, Boll. Un. Mat. Ital., 1 (1968) 517–526.
69. ______, The space of entire functions over certain nonarchimedean fields and its dual, Studia Math., 33 (1969) 251–256.
70. P. Robba, Quelques propriétés des éléments et fonctions analytiques au sens de Krasner, Bull. Soc. Math. France Mém., 39–40 (1974) 341–349.
71. ______, Fonctions analytiques sur les corps valués ultramétriques complets, Astérisque, 10 (1973) 109–220.

SUMS OF FUNCTIONS SATISFYING RECURSION RELATIONS

ELDON HANSEN
Lockheed Missiles & Space Co., P.O. Box 504, Sunnyvale, CA 94086

1. Introduction. Many of the functions studied in mathematics satisfy three-term recursion relations. This is true, for example, of orthogonal polynomials, Bessel functions, and Legendre functions. The recursion relations have been used in various ways to generate sums and infinite series involving the functions.

In this paper, we use the recursion relations in some very simple ways to generate some new sums. We consider three variations of a single method.

2. First Variation. Consider a recursion relation of the form

---

Eldon Hansen received his Ph.D. from Stanford in 1960 under G. Forsythe. He has held various positions at Lockheed Missiles and Space Co. since 1956, with time out for teaching at Stanford, Oxford, San Jose State, and Washington State universities. He is currently mislabeled (he says) as a Senior Staff Engineer. His research interest is in numerical analysis. —*Editors*

$$a_n f_n(z) + b_n f_{n+1}(z) + c_n f_{n+2}(z) = 0 \tag{2.1}$$

for some function $f_n(z)$ where n takes integer values.

The argument z of the functions will be the same throughout this paper. For brevity, we shall write f_n for $f_n(z)$. Note that the coefficients a_n, b_n, and c_n in (2.1) may also depend on z.

For the first variation, define

$$u_n = A_n f_n \tag{2.2}$$

where A_n is chosen so that

$$\frac{A_{n+1}}{A_n} = \frac{-b_n}{a_n}. \tag{2.3}$$

This assures that the sum of the first two terms in the recursion relation (2.1) is proportional to the forward difference $u_{n+1} - u_n$ of the function u_n. We then make use of the fact that a forward difference can be summed. Thus we consider

$$\sum_{n=k}^{m-1} (u_{n+1} - u_n) = \sum_{n=k}^{m-1} (A_{n+1} f_{n+1} - A_n f_n).$$

Because of pairwise cancellations in the sum in the left member, it becomes $u_m - u_k$. Substituting for A_{n+1} from (2.3),

$$u_m - u_k = -\sum_{n=k}^{m-1} (b_n f_{n+1} + a_n f_n)\frac{A_n}{a_n}$$

so that from (2.1) and (2.2)

$$A_m f_m - A_k f_k = \sum_{n=k}^{m-1} c_n f_{n+2}\frac{A_n}{a_n}. \tag{2.4}$$

This is the result we sought.

We now determine A_n. Using (2.3), we form the product

$$\prod_{n=0}^{r-1} \frac{A_{n+1}}{A_n} = \prod_{n=0}^{r-1} \frac{-b_n}{a_n}.$$

The left member simplifies because of pairwise cancellations and becomes A_r/A_0. Choosing $A_0 = 1$, we have

$$A_r = \sum_{n=0}^{r-1} \frac{-b_n}{a_n}. \tag{2.5}$$

This completes the derivation. The sum (2.4) is now known since A_r is given by (2.5).

3. Example. As an example, consider the recursion relation for the Hermite polynomials,

$$2(n+1)H_n - 2zH_{n+1} + H_{n+2} = 0. \tag{3.1}$$

Here $a_n = 2(n+1)$, $b_n = -2z$, and $c_n = 1$. From (2.5),

$$A_r = \frac{z^r}{r!}.$$

Substituting into (2.4),

$$\frac{z^m}{m!}H_m - \frac{z^k}{k!}H_k = \sum_{n=k}^{m-1} \frac{z^n}{2(n+1)!}H_{n+2}.$$

If we let $k = 0$, replace n by $n - 1$, and use the facts that $H_0 = 1$ and $H_1 = 2z$, we can rewrite this result as

$$\sum_{n=0}^{m} \frac{z^n}{n!} H_{n+1} = \frac{2}{m!} z^{m+1} H_m.$$

We believe this sum is new.

4. Second Variation. In the first variation of our method, we found a function u_n such that its forward difference was proportional to the sum of the first and second terms of the recursion relation (2.1). The second variation does the same thing for the sum of the second and third terms.

Define

$$v_n = B_n f_n \tag{4.1}$$

where B_n is chosen so that

$$\frac{B_{n+1}}{B_n} = -\frac{c_{n-1}}{b_{n-1}}. \tag{4.2}$$

As before, we consider the sum

$$\sum_{n=k}^{m-1} (v_{n+1} - v_n) = \sum_{n=k}^{m-1} (B_{n+1} f_{n+1} - B_n f_n).$$

Substituting for B_{n+1} from (4.2) and then using (2.1) and (4.1), we obtain

$$B_m f_m - B_k f_k = \sum_{n=k}^{m-1} a_{n-1} f_{n-1} \frac{B_n}{b_{n-1}}. \tag{4.3}$$

This is the desired sum.

Choosing $B_1 = 1$ and using (4.2), we find

$$B_r = \prod_{n=1}^{r-1} \frac{-c_{n-1}}{b_{n-1}}. \tag{4.4}$$

5. Example. Using the recursion (3.1) for the Hermite polynomials, we find from (4.4) that

$$B_r = (2z)^{1-r}.$$

Substituting into (4.3), we obtain

$$(2z)^{1-m} H_m - (2z)^{1-k} H_k = -\sum_{n=k}^{m-1} 2n(2z)^{-n} H_{n-1}.$$

This result is known. It is equivalent to equation (49.3.2) of [**1**].

6. Third Variation. In our third variation, we concentrate on the first and third terms of the recursion relation (2.1). Define

$$w_n = C_n f_n \tag{6.1}$$

and choose C_n so that

$$\frac{C_{n+2}}{C_n} = \frac{-c_n}{a_n}. \tag{6.2}$$

Then

$$\sum_{n=k}^{m-1} (w_{2n+2+s} - w_{2n+s}) = \sum_{n=k}^{m-1} (C_{2n+2+s} f_{2n+2+s} - C_{2n+s} f_{2n+s}).$$

Using (6.1), (6.2), and (2.1) in the same way as for the first two variations, we obtain

$$C_{2m+s}f_{2m+s} - C_{2k+s}f_{2k+s} = \sum_{n=k}^{m-1} \frac{b_{2n+s}f_{2n+s+1}C_{2n+s}}{a_{2n+s}}. \tag{6.3}$$

This is the desired sum.

We have introduced the variable s in order to be able to treat even and odd cases conveniently. We shall choose either $s = 0$ or $s = 1$.

To obtain C_{2m+s}, we form the product

$$\prod_{n=0}^{r-1} \frac{C_{2n+2+s}}{C_{2n+s}} = \prod_{n=0}^{r-1} \frac{-c_{2n+s}}{a_{2n+s}}.$$

Choosing $C_s = 1$,

$$C_{2r+s} = \prod_{n=0}^{r-1} \frac{-c_{2n+s}}{a_{2n+s}}. \tag{6.4}$$

7. Example. To illustrate the use of (6.3), we again use the recursion relation (3.1) for the Hermite polynomials. From (6.4),

$$C_{2r+s} = (-4)^{-r} \frac{\Gamma(s/2 + 1/2)}{\Gamma(r + s/2 + 1/2)}$$

so that (6.3) becomes (after simplification)

$$\sum_{n=k}^{m-1} \frac{1}{\Gamma(n + s/2 + 3/2)} (-4)^{-n} H_{2n+s+1}$$
$$= \frac{2}{z}\left[\frac{(-4)^{-k}}{\Gamma(k + s/2 + 1/2)} H_{2k+s} - \frac{(-4)^{-m}}{\Gamma(m + s/2 + 1/2)} H_{2m+s}\right].$$

For $s = k = 0$, this becomes

$$\sum_{n=0}^{m-1} \frac{(-4)^{-n}}{\Gamma(n + 3/2)} H_{2n+1} = \frac{2}{z}\left[\frac{1}{\Gamma(1/2)} - \frac{(-4)^{-m}}{\Gamma(m + 1/2)} H_{2m}\right].$$

This appears to be a new sum.

For $s = 1$ and $k = 0$, we obtain

$$\sum_{n=0}^{m-1} \frac{(-4)^{-n}}{(n + 1)!} H_{2n+2} = \frac{2}{z}\left[H_1 - \frac{(-4)^{-m}}{m!} H_{2m+1}\right].$$

This is equivalent to equation (49.4.3) of [**1**].

8. Conclusion. When our procedure is applied to the Chebyshev polynomials, known results are obtained. However, results using Laguerre, Gegenbauer, or Jacobi polynomials or Bessel or Legendre functions seem to be new.

Reference

1. E. R. Hansen, A Table of Series and Products, Prentice-Hall, Englewood Cliffs, N.J., 1975.

MISCELLANEA

64. "No calculus without calculation." ... The essence of mathematics is not its symbolism, but its methods of deduction.—R. B. Braithwaite, *Scientific Explanation*, Cambridge University Press, 1953, p. 366.

NORMED DIVISION DOMAINS

SOLOMON W. GOLOMB
Department of Mathematics, University of Southern California, Los Angeles, CA 90007

1. Introduction. Division is usually introduced as the inverse of multiplication, and the theory of divisibility, including prime and composite elements, is usually set in the context of integral domains. Much weaker assumptions suffice. Even a multiplicative semigroup has more structure than is required. If one geometric figure tiles another, this is an instance of division without corresponding multiplication. More generally, if β can be partitioned into "copies" (or "translates," or "versions") of α, we may say that α *divides* β. Abstractly, we define a *normed division domain* (NDD) to consist of a set S with a partial ordering, $<|$, and a norm N which maps S into the positive integers so that whenever $\alpha <| \beta$, with α and β in S, then $N(\alpha) | N(\beta)$, and we further require, for any η in S such that $N(\eta) = 1$, that $\eta <| \alpha$ for all α in S. (The norm is often a measure of what is being "used up" by the copies of α which exhaust β.)

This concept, as austere as it is, is sufficient to accommodate the notions of *units*, *primes*, and *composites*; of *least common multiples*, *greatest common divisors*, and *principal ideals*.

If the example of an NDD under consideration is in fact an integral domain (e.g., a ring of algebraic integers, or a polynomial ring over a field), then if $\alpha <| \beta$ there is a "quotient" γ such that also $\gamma <| \beta$ with α as the "quotient," and moreover β is uniquely recoverable as the "product" of α and γ. It is actually sufficient to have a multiplicative semigroup with a norm function to have effective symmetry between "divisors" and "quotients," and the recoverability of the "dividend" as the unique "product" of divisor and quotient.

In the structure theory of finite groups, there is *duality* (but not *symmetry*) between "divisors" and "quotients" (*i.e.*, *kernels* and *factor groups*). That is, there are two *different* partial orderings with the same set $S = \{\text{all finite groups}\}$ and norm $N(G) =$ order of G, with both NDD's having the same *primes* (i.e., the simple groups). However, given both the "divisor" and the "quotient," it is not in general possible to specify the "dividend" uniquely. Thus, even though all the finite simple groups have been determined, more work is needed to characterize the set of all finite groups.

If S is the set of all finite graphs, there are two different *norms* which suggest themselves: $E(G) =$ the number of *edges* of the graph G, and $V(G) =$ the number of *nodes* (or *vertices*) of G. There is a corresponding notion of division appropriate to each of these norms. The V-norm leads to a divisibility theory which includes, among its applications, the notion of *polyomino division* (where α divides β if copies of α can be used to tile β). The E-norm gives rise to a divisibility theory which encompasses a surprising variety of classical combinatorial problems, including the existence of Steiner Triple Systems, finite projective planes, and other balanced incomplete block designs; and the conjecture that all trees are graceful.

2. Normed Division Domains

DEFINITION 1. Let S be a partially ordered set, with a partial ordering relation "$<|$", and a norm N which maps S into the positive integers such that if $\alpha <| \beta$ for α and β in S, then $N(\alpha) | N(\beta)$, and if $N(\eta) = 1$ for η in S, then $\eta <| \alpha$ for all $\alpha \in S$. Then $D = (S, <|, N)$ is called a *normed division domain*, abbreviated NDD.

DEFINITION 2. If $N(\eta) = 1$ for an element η in an NDD, then η is called a *unit* in that NDD.

Notes. 1. Our convention is that the partial ordering "$<|$" is reflexive as well as transitive: $\alpha <| \alpha$ for all α in S; if $\alpha <| \beta$ and $\beta <| \gamma$, then $\alpha <| \gamma$.

A biographical note about the author appears in this MONTHLY, 85 (1978) 734.

2. We read "$\alpha <| \beta$" as "α divides β"; and "$m|n$" means "m divides n" in the ring of ordinary integers.

3. In every example of an NDD we will consider, there exists at least one unit.

DEFINITION 3. Let $D = (S, <|, N)$ be an NDD. Then π in S is a *prime* of D if $N(\pi) > 1$ and there is no α in S with $\alpha <| \pi$ and $1 < N(\alpha) < N(\pi)$.

THEOREM 1. *Let $D = (S, <|, N)$ be an* NDD. *A sufficient condition for an element π in S to be a prime of D is that $N(\pi)$ be a rational prime.*

Proof. If $N(\pi)$ is prime and $\alpha <| \pi$, then since $N(\alpha)|N(\pi)$ either $N(\alpha) = 1$ or $N(\alpha) = N(\pi)$. ∎

THEOREM 2. *Let $D = (S, <|, N)$ be an* NDD. *An element π in S with $N(\pi) > 1$ is a prime of D if and only if there is no element ρ in S which is a prime of D such that $\rho <| \pi$ and $N(\rho) < N(\pi)$.*

Proof. If there *is* a prime ρ of D with $\rho <| \pi$ and $N(\rho) < N(\pi)$, then clearly π is *not* a prime of D. Conversely, if $N(\pi) > 1$ and π is not a prime of D, we must show that there exists a "proper prime divisor" of π. For this, we will use induction on $N(\pi)$. If $N(\pi) = 2$, the smallest possible norm for a prime, then in fact π must be prime by Theorem 1. Assume the result for all elements of norm $\leqslant m$, where $m \geqslant 2$, and consider an element π with $N(\pi) = m + 1$. By Definition 3, either π is prime, or it has a divisor ρ with $2 \leqslant N(\rho) \leqslant m$. By the inductive assumption, either ρ is prime, in which case there is nothing further to prove, or ρ has a *prime* divisor σ, with $2 \leqslant N(\sigma) < N(\rho) \leqslant m$. In this case, since $\sigma <| \rho$ and $\rho <| \pi$ and the relation "$<|$" is transitive, $\sigma <| \pi$ and σ is a prime divisor of π with $1 < N(\sigma) < N(\pi)$.

The next two results describe the preservation of primes under restriction or refinement of an NDD. The proofs consist of straightforward verification, and are omitted.

THEOREM 3. *Let $D = (S, <|, N)$ be an* NDD. *Let T be a subset of S, and let $D' = (T, <|, N)$ be the result of restricting "$<|$" and "N" from S to T. Then D' is an* NDD, *and every prime of D which is an element of T is a prime of D'.*

THEOREM 4. *Let $D = (S, <|, N)$ be an* NDD. *Let "$\ll|$" be a partial ordering on S such that $\alpha \ll| \beta$ whenever $\alpha <| \beta$, but not, in general, conversely, and such that $N(\alpha)|N(\beta)$ whenever $\alpha \ll| \beta$. Then $D^* = (S, \ll|, N)$ is an* NDD, *and if π in S is a prime of D^*, then π is a prime of D.*

DEFINITION 4. Let $D = (S, <|, N)$ be an NDD. For $\alpha \in S$, we define (α), the *principal ideal of* α, to be the set of all $\sigma \in S$ such that $\alpha <| \sigma$. If $\eta \in S$ with $N(\eta) = 1$, then $(\eta) = S$ is the *unit ideal.*

DEFINITION 5. With $D = (S, <|, N)$ and $\alpha \in S$, $\beta \in S$, we define the set $M = M(\alpha,\beta)$ of *common multiples* of α and β to be $(\alpha) \cap (\beta)$. If $M \neq \varnothing$, then M has a nonempty subset $\mathrm{LCM}(\alpha,\beta)$ of elements of least norm, called the *least common multiple*(s) of α and β. (This least norm must be a common multiple of $N(\alpha)$ and $N(\beta)$, but not necessarily their *least* common multiple.)

Note. One can show by examples that the property "α and β have a (least) common multiple" is reflexive and symmetric, but not necessarily transitive.

DEFINITION 6. With $D = (S, <|, N)$ and $\alpha \in S$, $\beta \in S$, we define the set $C = C(\alpha,\beta)$ of *common divisors* of α and β to consist of those elements $\gamma \in S$ such that $\alpha \in (\gamma)$ and $\beta \in (\gamma)$. (If there exists $\eta \in S$ with $N(\eta) = 1$, then $\eta \in C(\alpha,\beta)$ for all $\alpha \in S$ and $\beta \in S$.) If $C \neq \varnothing$, then C has a nonempty subset $\mathrm{GCD}(\alpha,\beta)$ of elements of greatest norm, called the *greatest common divisor*(s) of α and β. (This greatest norm must be a common divisor of $N(\alpha)$ and $N(\beta)$, but not necessarily their *greatest* common divisor.)

Another notion from classical algebraic number theory which it is useful to generalize to arbitrary NDD's is that of "associates."

DEFINITION 7. Let $D = (S, <|, N)$ be an NDD. Let S' be the set of equivalence classes, or "orbits," in S established by a norm-preserving equivalence relation. We define a new NDD by $D' = (S', <|, N)$, where if $\sigma \in S'$ and $\tau \in S'$, then $\sigma <| \tau$ if and only if there exist elements s and t in S with $s \in \sigma$ and $t \in \tau$, such that $s <| t$ in D. Elements of S belonging to the same orbit in S' are called *associates* (relative to division in D').

In specific instances of NDD's, the orbits are likely to arise from some systematic group of transformations on the elements of S. In algebraic number theory, this is explicitly the group of units. In fact, it is *consistent* with division in any NDD to place all units into a single orbit. However, Definition 7 allows considerable latitude in the choice of associates in an arbitrary NDD.

3. Some Examples. A. Suppose that S is the set of nonzero Gaussian integers, with the customary notion of divisibility, and as usual $N(a + bi) = a^2 + b^2$.

Let T be the subset of S consisting of the ordinary integers, with standard divisibility and the same norm: $N(a) = N(a + 0 \cdot i) = a^2$. Theorem 3 asserts that every Gaussian prime which is an ordinary integer is a prime of T. The converse (that every prime of T is a prime of S) is of course false.

It is conventional to define *associates* in S by the rule that the orbit of $z = a + ib$ in S consists of $\{z, iz, -z, -iz\}$, corresponding to transformations by the group C_4 of units in S. It would also be *consistent* with our Definition 7 to define *associates* as members of a larger orbit, $\{z, iz, -z, -iz, \bar{z}, i\bar{z}, -\bar{z}, -i\bar{z}\}$, generated by the group D_4 consisting of complex conjugation as well as multiplication by units. The corresponding divisibility theory for the Gaussian integers differs only trivially from the usual one.

B. Let S be the set of all (abstract) finite groups, and for $G \in S$, define $N(G) =$ order of G. We consider three *different* partial orderings on S with respect to this norm, each giving rise to a different NDD.

(i) $D_0 = (S, |, N)$ where $H|G$ if H is (isomorphic to) a subgroup of G. In this NDD, the only primes are the cyclic groups of prime order, i.e., those elements guaranteed to be prime by Theorem 1.

(ii) $D_1 = (S, <|, N)$ where $H <| G$ if H is (isomorphic to) a *normal* subgroup of G, or (since this alone would not be transitive) if there is a *normal series* from H to G, *i.e.*, groups $H_1, H_2, \ldots, H_k$ with $H \lhd H_1 \lhd H_2 \lhd \cdots \lhd H_k \lhd G$, where "$H \lhd G$" means "$H$ is a normal subgroup of G." (Some authors say, in this case, that H is *subnormal* in G.)

In D_1, the primes are precisely the finite simple groups. By Theorem 4, these must include the primes of D_0, namely, the cyclic groups of prime order, as a subset.

Note that $D^* = (S, \lhd, N)$ is *not* an NDD, because "$\lhd$" is not transitive. Also, Theorem 2 fails in D^*: a finite group may fail to be simple even though it has no *simple* normal subgroup. (The symmetric group S_4 is a case in point.)

(iii) $D_2 = (S, |>, N)$ where $H |> G$ if there is a homomorphism from G to H. In D_2 as in D_1 the primes are the finite simple groups, but the NDD's are different. Thus "$C_3 <| S_3$" and "$C_2 |> S_3$" are both true, while "$C_2 <| S_3$" and "$C_3 |> S_3$" are both false. In ordinary group theory notation, $S_3/C_3 \cong C_2$, with asymmetry between the *divisor* (*kernel*) C_3 and the *quotient* (*factor group*) C_2.

There is an obvious *duality* between $<|$ and $|>$, and the lattice structures D_1 and D_2 which they impose on the set of all finite groups. Lagrange's Theorem, that if H is a subgroup of G then

$N(H)|N(G)$, is clearly tailored to the requirements for an NDD. It is also noteworthy that our "$H <|G$" goes back to the very beginnings of group theory, in the form of "normal series," related to "composition series" and the *solvability* of both finite groups and algebraic equations.

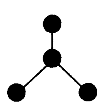
The five groups of order 8 provide ample illustration that given any two of "dividend," "divisor," and "quotient" in the equation $G/H \cong J$, this may be insufficient to specify the third member uniquely. Thus, if $G/C_2 \cong V_4$, where $V_4 = C_2 \times C_2$ is Klein's "fours group," there are *four* possible choices for G (either $C_2 \times C_2 \times C_2$ or $C_2 \times C_4$ or D_4 or the quaternion group Q). From $D_4/H \cong C_2$, we do not know whether $H = C_4$ or $H = V_4$; and if we let $B = C_2 \times C_4$, then $B/C_2 \cong J$ allows both $J = C_4$ and $J = V_4$. This is a far cry from the situation which obtains even in multiplicative semigroups, where knowing the "factors" uniquely specifies the "product."

C. The set S of all Hadamard matrices forms a semigroup under the operation "$*$" of Kronecker product. If $H \in S$ is $n \times n$, we may take $N(H) = n$, since $H_1 * H_2$ is $mn \times mn$ if H_1 is $m \times m$ and H_2 is $n \times n$. The Kronecker product is noncommutative, so that one consistent definition of $H' <|H$ would be that there exists a Hadamard matrix H'' with $H' * H'' = H$. A dual divisibility notion would be defined by $H''|> H$ if there exists a Hadamard matrix H' with $H' * H'' = H$. (These two divisibility notions obviously occur in any noncommutative semigroup.)

From the standpoint of research on the existence of Hadamard matrices, it is useful to define *associates* among these matrices by the rule that H_2 is an associate of H_1 if H_2 can be obtained from H_1 by some combination of: permutation of rows, permutation of columns, complementation of rows, complementation of columns, and transpose. (Note that now $H_1 * H_2$ and $H_2 * H_1$ are associates.) The primes in this system are then the irreducible building blocks (relative to Kronecker products and the associate operations) in the construction and classification of all Hadamard matrices. The identification of all *prime* Hadamard matrices would complete the classification of all Hadamard matrices much more straightforwardly than the identification of all simple groups completes the classification of all finite groups. On the other hand, the finite simple groups have been completely classified, while the knowledge of all prime Hadamard matrices is still in a rather rudimentary state.

D. The set Γ of all (finite, connected, undirected) graphs has a norm $V(G)$ = number of nodes (vertices) of G. We define division relative to this norm by: $H <|G$ if nonoverlapping isomorphic copies of H can be placed so as to cover all the *nodes* (but not necessarily all the edges, some of which may be left uncovered) of G. We require the $V(G)/V(H)$ copies of H which "cover" G to be disjoint subgraphs of G. (In a more restricted notion of graph division, we would require each copy of H to be the induced subgraph on the $V(H)$ nodes of G which it covers. The examples given in this paper are not affected by the adoption or removal of this restriction.)

In this NDD, $D = (\Gamma, <|, V)$, the graph consisting of a single node (and no edges) is the unique unit. Any graph with a prime number of nodes must be prime, by Theorem 1. The only prime with $V(G) = 4$ is the tree



This is prime as an instance of the following remarkable theorem.

THEOREM 5. If G is a bicolorable graph with $V(G) = 2^k$ for $k \geqslant 1$, and the 2-coloring of G results in an odd number of nodes of each color, then G is prime in $D = (\Gamma, <|, V)$.

Note. A *bicolorable graph* is one whose nodes may be colored in two colors such that no two nodes of the same color are joined by an edge. All trees and $2n$-gons, and in fact all bipartite graphs, are bicolorable. ("Bicolorable" and "bipartite" are equivalent properties.)

Proof. Let G satisfy the hypotheses of the theorem. By Theorem 2, G is prime unless it has a

proper prime divisor H, and by Theorem 1, we must have $V(H) = 2^j$ with $1 \leqslant j < k$. We use induction on k. At $k = 1$, $V(G) = 2$ is prime, so G, which splits one-one in the coloring, is prime. Assume the result for all $1 \leqslant k < k_0$, and consider G with $V(G) = 2^{k_0}$. If a bicolorable graph G is divisible by the (unique) graph K_2 with two nodes, then G will have *equally many* nodes, namely, 2^{k_0-1}, of each color, and since $k_0 > 1$, this is an *even number* of each. (Each copy of K_2 in G must cover one node of each color!) Thus G cannot be divisible by K_2. Suppose G is divisible by prime H, with $V(H) = 2^j$, $1 \leqslant j < k_0$. Then the number of copies of H which cover G is 2^{k_0-j}, an *even* number. Each copy of H splits odd-odd (by the inductive assumption), but the sum of an even number of such splits must come out even-even, contradicting the hypothesis about G. ∎

Note. The proof works equally well without restricting H to be a *prime* divisor. An even number of even-even splits also sum to even-even.

Among the twenty-three trees with 8 nodes, ten split 5–3 and one splits 7–1. These are prime by Theorem 5. It also turns out that two of the three trees which split 6–2 are prime, as are two of the nine trees which split 4–4. The remaining eight trees are composite.

The ideas contained in the proof of Theorem 5 are sufficient to obtain the following generalization.

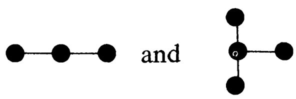
THEOREM 6. *Let $V(G) = 2^k m$ with m odd, where G is bicolorable and splits odd-odd in the 2-coloring of its nodes. If $H <| G$, then either*

(i) *$V(H)$ is odd,*

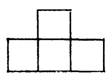
or

(ii) *$V(H)$ is even, but $V(G)/V(H)$ is odd, and H splits odd-odd in the 2-coloring of its nodes.*

As an application of Theorem 6, if a tree with 12 nodes splits odd-odd in the 2-coloring of its nodes, its only possible proper prime factors are



E. The set P_2 of all polyominoes in the plane, with $N(\alpha) = n$ if α is an n-omino, and $\alpha <| \beta$ if copies of α can be used to tile β, forms an NDD, $D = (P_2, <|, N)$. This is really an application of the previous example, since we may regard an n-omino as a graph with n nodes (the centers of the unit squares), where two nodes are connected by an edge if the corresponding unit squares are adjacent. All polyominoes are bicolorable (the checkerboard coloring!), so Theorem 5 applies. The only prime tetromino is



(The other four tetrominoes are divisible by the domino.) Of the 369 octominoes, 145 are prime, and of these, 136 split 5–3 in the checkerboard coloring.

Separate papers, in preparation, will describe in considerably greater detail the divisibility theory of: (i) graphs with respect to the V-norm, (ii) polyominoes in two or more dimensions, and (iii) other tiling figures such as those formed from congruent equilateral triangles, as well as (iv) graphs with respect to the E-norm. Note that the divisibility theory for tiling by one-dimensional polyominoes (line segments of integer length) reduces to the ordinary divisibility theory of the positive integers, in basically the same geometric representation as that visualized by Euclid.

F. Let Γ again be the set of all finite graphs, and define the E-norm by $E(G)$ = number of edges of the graph G. Relative to this norm, we define division by:

$H \rhd G$ if isomorphic copies of H can be used to cover all the *edges* of G, each edge exactly once, where it is allowed to cover the nodes of G more than once each. Then $(\Gamma, \rhd, E)$ is an NDD.

Notes. 1. There is an analogy between the relationship of edge division to vertex division, and the relationship of Euler circuits to Hamiltonian circuits. Edge division and Euler circuits require covering each edge of the graph exactly once, with the vertices covered at least once each. Vertex division and Hamiltonian circuits require covering each node of the graph exactly once, with the edges covered at most once each.

2. The definition of $H \rhd G$ requires that the copies of H used to cover G remain isomorphic to H. Thus if H is the *tree*,

,

it cannot be used to cover a *triangle* in G as part of the division of G by H.

Let K_n be the complete graph on n nodes. We have $E(K_n) = n(n-1)/2$, and in particular $E(K_2) = 1$, and we observe that K_2, which consists of a single edge (with its two terminal nodes) is the unique *unit* of the NDD, $(\Gamma, \rhd, E)$.

The question of determining the divisors of K_n in the E-norm already contains many difficult combinatorial problems (some solved, most unsolved) as special cases. The following two groupings of such problems are illustrative of the richness of this theory. (I am indebted to Herbert Taylor for identifying $D = (\Gamma, \rhd, E)$ as an important instance of an NDD, and for providing virtually all of the examples which follow.)

(i) When does K_a divide K_b in the E-norm? Two necessary conditions for $K_a \rhd K_b$ are the *valence* condition: $(a-1)|(b-1)$, and the *norm* condition:

$$\frac{a(a-1)}{2} \,\Big|\, \frac{b(b-1)}{2}.$$

For what values of n does $K_3 \rhd K_n$? This is precisely the question of the existence of Steiner Triple Systems, for which the two necessary conditions

$$\left(2|(n-1) \text{ and } 3\Big|\frac{n(n-1)}{2}\right)$$

are also sufficient. That is, $K_3 \rhd K_n$ iff $n \equiv 1$ or $3 \pmod 6$.

The existence of a finite projective plane of order n is equivalent to $K_{n+1} \rhd K_{n^2+n+1}$. Thus we know that $K_7 \rhd K_{43}$ is false (there is no plane of order 6, even though the two necessary conditions are satisfied), and whether $K_{11} \rhd K_{111}$ is still an open question (whether there exists a plane of order 10). In general, $K_m \rhd K_n$ iff there exists a balanced incomplete block design (BIBD) with the parameters $v = n$, $k = m$, b, r, $\lambda = 1$, where as usual $vr = kb$ and $\binom{v}{2} = b\binom{k}{2}$.

(ii) Let T be a tree with $E(T) = e$. A conjecture of Ringel **[1]** can be stated in the form: $T \rhd K_{2e+1}$, for all T with $E(T) = e$.

We next define a more restrictive notion of divisibility of the graphs K_n relative to the E-norm, called *cyclic divisibility*, and denoted by "Φ", as follows: Represent K_n as a regular n-gon with all the diagonals drawn in. Then $G \Phi K_n$ means that G can be placed on this diagram of K_n so that the n rotations of G around the center of K_n by the multiples of $2\pi/n$ accomplish the division (in the "$\rhd$" sense) of K_n by G. Suppose T is a tree with $E(T) = e$. Then it turns out that $T \Phi K_{2e+1}$ iff T is *graceful modulo* $2e+1$. If $T \Phi K_{2e+1}$ in such a way that one copy (and hence all copies) of T in the cyclic division of K_{2e+1} subtends an arc of less than 180° on the circumscribed circle of K_{2e+1}, it turns out that this is equivalent to T being graceful. It is widely conjectured [2] that all trees are graceful. Thus we see a hierarchy of tree-conjectures:

(a) Ringel's conjecture: $T \text{þ} K_{2e+1}$ whenever $E(T) = e$.

(b) Every tree with e edges is graceful modulo $2e + 1$:

$$T \Phi K_{2e+1} \text{ whenever } E(T) = e.$$

(c) All trees are graceful: $T \phi K_{2e+1}$ with the "180° restriction" whenever $E(T) = e$.

Since the strongest of these three conjectures, viz. (c), is believed to be true, it is inefficient to look for counterexamples to (a) or (b). However, examples of graphs H, which are not trees, have been found which demonstrate that, with $E(H) = e$, the corresponding three situations, namely

(a) $H \text{þ} K_{2e+1}$

(b) $H \phi K_{2e+1}$

and

(c) H is graceful

are all logically distinct.

Analogous results have been obtained regarding the divisibility of the *complete bipartite graphs*, $K_{m,n}$, relative to the E-norm.

This research was supported in part by the United States Navy, Office of Naval Research, under Contract No. N00014-80-C-0745.

References

1. G. Ringel, Problem 25, Theory of graphs and its applications, Proc. Int. Symp., Smolenice, June, 1963, Czech. Acad. Sci., Prague, 1964, p. 162.

2. A. Rosa, On certain valuations of the vertices of a graph, in Theory of Graphs, Gordon and Breach, New York, 1967, pp. 349-355.

EIGENVALUES OF THE LAPLACIAN AND THE HEAT EQUATION

JOZEF DODZIUK

University of Oxford Mathematical Institute, Oxford OXI 3LB, England

Heat equation methods have been very successful in applications involving spectra of elliptic operators. Recent results along these lines have been obtained in very sophisticated contexts (cf. [**BGM**], [**McKS**], [**P**], [**ABP**]). The purpose of this article is to show how one of the oldest results about eigenvalues of the Laplacian, the asymptotic formula of Hermann Weyl (cf. (3.1) below), follows from basic facts about the heat equation, and to develop the main properties of the eigenvalues and eigenfunctions from the same point of view. There is not much new in the proofs we present. As a matter of fact, the basic facts about the heat equation used in this paper had been known (cf. [**L**]) at the time H. Weyl gave the proof of his formula [**W1**]. The original proof did not use the heat equation, whose significance became apparent much later in the work of Minakshisundaram [**M1**], [**M2**], Minakshisundaram and Pleijel [**MP**], Milgram and Rosenbloom [**MR**], and Kac [**K**].

Jozef Dodziuk received his Ph.D. in 1973 at Columbia University under the guidance of Lipman Bers and Howard Garland. He taught at MIT and at the University of Pennsylvania before going to Queens College of CUNY, where he is an Associate Professor of Mathematics. He held visiting positions at Bonn University, University of Oxford, and Universidad Tecnica del Estado in Santiago, Chile, and is now an Alfred P. Sloan Research Fellow. His main mathematical interests are relations between geometry, topology, and analysis on Riemannian manifolds.—*Editors*

Historically the eigenvalues and eigenfunctions of the Laplacian, i.e., solutions of the equation $\Delta\phi + \lambda\phi = 0$, appeared in connection with equations of mathematical physics such as the wave equation and the heat equation. The knowledge of the eigenfunctions and eigenvalues enables one to write the solutions of these equations explicitly in the form of infinite series. Of course this method applies only for very special domains (e.g., intervals on the line or squares in the plane) for which the eigenvalues and eigenfunctions are known explicitly. For most domains (e.g., triangles in the plane) the eigenvalues and eigenfunctions are not known. However, one can turn around and use the heat equation and its properties for a qualitative study of eigenvalues and eigenfunctions. This procedure is very general and can be applied in a variety of situations. We shall illustrate it in the simplest case of the Laplacian on a bounded domain in $\mathbb{R}^n$ with Dirichlet boundary conditions. We feel that this will serve as a useful introduction to heat equation methods.

The paper consists of three sections. In the first, we discuss the maximum principle for the heat equation and the Green's function for the heat equation in a bounded domain in $\mathbb{R}^n$ with Dirichlet boundary conditions. No proofs are given but we provide references and discuss the physical motivation. In the second section we use the Green's function and the uniqueness of solutions of the heat equation to prove the existence and completeness of the eigenfunctions of the Laplacian. We also prove the formula expressing the Green's function in terms of eigenfunctions and eigenvalues. Finally, in the third section we use the maximum principle to obtain the estimates of the Green's function which yield Weyl's formula. The technique used here is very special. The maximum principle is not applicable to problems with other boundary conditions or to systems. We wish to emphasize that estimates of the Green's function yield information about eigenvalues (and eigenfunctions). How these estimates are obtained depends on the context. In the case under consideration the maximum principle is a very convenient tool, but in other contexts different methods of deriving estimates have to be used. However, after such estimates are obtained, they can be used to derive analogs of Weyl's formula in the same way we prove formula (3.1).

This article grew out of seminar lectures given at the University of Pennsylvania in the fall of 1979 and at the Universidad Tecnica del Estado in Santiago, Chile, in June 1980. The author is grateful to organizers and participants of these seminars for discussions and comments.

1. The Heat Equation. The physical interpretation of our mathematical problem is the following. We study the heat flow in a domain D whose boundary is kept at temperature absolute zero. If $u(x, t)$ denotes the temperature at $x \in \bar{D}$ at time $t > 0$ and $u_0(x)$ is the temperature at x at time $t = 0$, then $u(x, t)$ satisfies

$$\begin{aligned} &\Delta u(x,t) = \frac{\partial u}{\partial t}(x,t) && \text{for} \quad x \in D,\ t > 0 \\ &u(x,t) = 0 && \text{for} \quad x \in \partial D,\ t \geqslant 0 \\ &u(x,0) = u_0(x) && \text{for} \quad x \in D. \end{aligned} \tag{1.1}$$

Here

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2},$$

and we require that D be an open, bounded set with a smooth (say C^∞) boundary ∂D. If we accept the physical interpretation, it is intuitively obvious that, given a function u_0 continuous in $\bar{D}$ and vanishing on ∂D, a unique solution of the problem (1.1) exists. We discuss the existence and uniqueness of solutions in more detail.

Uniqueness follows, for example, from the maximum principle stated below.

THEOREM 1.2. *Suppose $u(x, t)$ is continuous in $x \in \bar{D}$, $t \in [0, T]$, u is C^2 in the space variables $x = (x_1, \dots, x_n)$, C^1 in t, and satisfies the heat equation $\Delta u = \partial u/\partial t$ in $D \times (0, T)$. Then*

$$\max_{(x,t)\in\bar{D}\times[0,T]} u(x,t) = \max_{(x,t)\in(\bar{D}\times\{0\})\cup(\partial D\times(0,T])} u(x,t).$$

Physically, since the heat flows from points of higher temperature to those of lower temperature, the temperature cannot attain a maximum at an interior point unless $t = 0$.

The proof of this weak version of the maximum principle is very easy. It can be found in [**PW**, p. 162].

The set $\Gamma = (\bar{D} \times \{0\} \cup (\partial D \times [0,T])$ is called the normal boundary of the region $D \times (0,T)$ in $\mathbb{R}^{n+1}$. We remark that the maximum principle applied to $-u$ implies that under the assumptions of Theorem 1.2 the minimum of u is attained on the normal boundary as well.

From the maximum principle the uniqueness of the solutions of (1.1) follows easily. Suppose u and u^1 are two solutions. Then $v = u - u^1$ satisfies (1.1) with $u_0 \equiv 0$. Thus the function v satisfies the heat equation on $D \times (0,T)$ for every $T > 0$ and vanishes on the normal boundary. It follows that $\max_{\bar{D}\times[0,T]} v = \min_{\bar{D}\times[0,T]} v = 0$, i.e., $v = u - u^1 \equiv 0$.

To prove existence we may use the Green's function for (1.1). We describe it first in physical terms. Suppose D is at temperature zero initially and a unit of heat is introduced at $x \in D$ at time $t = 0$. Let $p(x,y,t)$ denote the resulting temperature at $x \in \bar{D}$ at time $t > 0$. We expect the function $p(x,y,t)$ to have the following properties.

$$\begin{aligned}
&(a)\ p(x,y,t) \geqslant 0,\quad p(x,y,t) = p(y,x,t),\quad p(x,y,t) = 0 \quad \text{if } x \in \partial D.\\
&(b)\ \frac{\partial}{\partial t} p(x,y,t) = \Delta_x p(x,y,t) \quad \text{for } (x,y,t) \in D \times D \times (0,\infty).\\
&(c)\ p(x,y,t+s) = \int_D p(x,z,t)p(z,y,s)\,dz.
\end{aligned} \tag{1.3}$$

(c) may be somewhat mysterious at this point. Before we try to make it plausible we observe that the function $p(x,y,t)$ can be used to construct solutions of (1.1). The temperature $u(x,t)$ at $x \in D$ at time $t > 0$ is the sum (integral) of effects due to individual points; i.e.,

$$u(x,t) = \int_D p(x,y,t)u_0(y)\,dy. \tag{1.4}$$

The formula (c) above expresses the fact that the temperature distribution after $t + s$ seconds is equal to the temperature after t seconds, if the initial values are the temperatures after s seconds. We now give a formal

DEFINITION 1.5. *A function $p(x,y,t)$ continuous on $\bar{D}\times\bar{D}\times(0,\infty)$, C^2 in x and C^1 in t is called the Green's function for the problem* (1.1) *if*

(a) $$\frac{\partial p}{\partial t} = \Delta_x p \quad on\ D \times D \times (0,\infty),\quad p(x,y,t) = 0$$

whenever $x \in \partial D$ and

(b) $$\lim_{t\to 0^+} \int_D p(x,y,t)u_0(y)\,dy = u_0(x)$$

uniformly for every function u_0 continuous on $\bar{D}$ and vanishing on ∂D.

If such a function exists, differentiation under the integral and (b) show that the formula (1.4) gives the solution of (1.1). It is also fairly easy to see that the Green's function is unique (this follows from uniqueness of solutions of (1.1)) and has all the properties stated in (1.3).

Existence of the Green's function was first proved in [**L**]. The construction is rather lengthy and technical, although it can be carried out by elementary means. We make some remarks about the construction before giving more modern references.

The function

$$k(x, y, t) = (4\pi t)^{-n/2}\exp\left(-\frac{|x-y|^2}{4t}\right) \tag{1.6}$$

plays the role of the Green's function for the whole space, i.e., it gives the temperature at $x \in \mathbb{R}^n$ at time $t > 0$ due to the unit of heat at time $t = 0$ at y if the body conducting heat fills the whole space. Calculations and easy estimates yield the following properties

(a) $k(x, y, t) \geqslant 0, \quad k(x, y, t) = k(y, x, t)$

(b) $\Delta_x k(x, y, t) = \frac{\partial}{\partial t}k(x, y, t) \quad \text{for } (x, y, t) \quad \mathbb{R}^n \times \mathbb{R}^n \times (0, \infty)$

(c) $\int_{\mathbb{R}^n} k(x, y, t)\,dy = 1 \quad \text{for every } t > 0, \quad x \in \mathbb{R}^n$

(d) $\int_{\mathbb{R}^n} k(x, z, t)k(z, y, s)\,dz = k(x, y, t + s)$

(e) $\lim_{t\to 0^+} \int_{\mathbb{R}^n} k(x, y, t)u_0(y)\,dy = u_0(x)$

uniformly, for every bounded and continuous function u_0 on $\mathbb{R}^n$.

We remark that the formula

$$u(x, t) = \int_{\mathbb{R}^n} k(x, y, t)\, u_0(y)\, dy$$

yields a solution of the initial value problem

$$\begin{aligned} \Delta u &= \frac{\partial u}{\partial t} && \text{on } \mathbb{R}^n \times (0, \infty) \\ u(x,0) &= u_0(x) && \text{for } x \in R^n \end{aligned}$$

whenever u is bounded and continuous.

Since $k(x, y, t)$ is explicitly known, we use it as the departure point for the construction of the Green's function. We expect that for small t the heat flow in D behaves similarly to the heat flow in $\mathbb{R}^n$, at least far from the boundary ∂D. This is the famous "principle of not feeling the boundary" (cf. [**K**]), and leads to conjecturing that

$$p(x, y, t) = k(x, y, t) - g(x, y, t), \tag{1.7}$$

where the function $g(x, y, t)$ is continuous in $\bar{D} \times \bar{D} \times (0, \infty)$, is C^2 in x and C^1 in t and satisfies

$$\begin{aligned} \Delta_x g(x, y, t) &= \frac{\partial}{\partial t} g(x, y, t) && \text{for } (x, y, t) \in D \times D \times (0, \infty) \\ g(x, y, t) &= k(x, y, t) && \text{for } x \in \partial D, \quad y \in D, \quad t > 0 \\ \lim_{t\to 0^+} g(x, y, t) &= 0 && \text{for } x \in \bar{D}, \quad y \in D. \end{aligned} \tag{1.8}$$

Thus, for every fixed $y \in D$, the function $v(x, t) = g(x, y, t)$ is a solution of the following initial-boundary value problem

$$\begin{aligned} \Delta v &= \frac{\partial v}{\partial t} && \text{in } D \times (0, \infty) \\ v(x, t) &= k(x, y, t) && x \in \partial D, \quad t > 0 \\ v(x,0) &= 0 && \text{for } x \in \bar{D}. \end{aligned} \tag{1.8$'$}$$

Now the construction of $p(x, y, t)$ may proceed as follows. We first solve infinitely many (one for every $y \in D$) problems (1.8′), piece the solutions together to obtain $g(x, y, t)$ satisfying (1.8)

and define the Green's function by (1.7). The initial boundary value problems of the type (1.8′) are discussed very thoroughly in [**F**] for a very large class of equations which contains the heat equation. Another excellent reference is [**LSU**] which also treats a very general theory.

2. Eigenvalues and Eigenfunctions. Recall that a function $\phi \not\equiv 0$ is called an eigenfunction of the Laplacian in D with Dirichlet boundary conditions if ϕ is C^2 in D, ϕ is continuous in $\overline{D}$ and satisfies

$$\begin{aligned} \Delta\phi + \lambda\phi &= 0 \\ \phi \mid \partial D &= 0 \quad . \end{aligned} \tag{2.1}$$

for some constant λ. One says that ϕ belongs to the eigenvalue λ.

The following theorem explains the relation between eigenvalues and eigenfunctions on one hand and the heat equation on the other. The proof was suggested by [**MR**].

Let $L^2(D)$ denote the Hilbert space of complex valued square-integrable functions on D with the inner product given by $(f, g) = \int_D f(x)\overline{g(x)}dx$.

THEOREM 2.2. *There exists a complete orthonormal basis* $\{\phi_i\}_{i=1}^{\infty}$ *of* $L^2(D)$ *consisting of eigenfunctions of* Δ. *If* λ_i *is the eigenvalue belonging to* ϕ_i, *then* $\lambda_i > 0$ *and* $\lim_{i\to\infty}\lambda_i = \infty$. *Moreover*

$$p(x, y, t) = \Sigma e^{-\lambda_i t}\phi_i(x)\overline{\phi_i(y)}$$

where the convergence is uniform on $\overline{D} \times \overline{D} \times [\varepsilon, \infty)$ *for every* $\varepsilon > 0$.

The proof uses three well-known results about integral equations (cf. [**H**]), which we state below.

THEOREM 2.3 [**H**, Ch. 3, Theorem 6]. *Let* $K(x, y)$ *be complex valued and continuous on* $\overline{D} \times \overline{D}$ *and* $K(x, y) = \overline{K(y, x)}$. *Then the operator* $A\colon L^2(D) \to L^2(D)$ *given by*

$$Af(x) = \int_D K(x, y)f(y)\, dy$$

is a compact self-adjoint operator.

THEOREM 2.4 (Spectral theorem for compact self-adjoint operators, [**H**, Ch. 3, Theorem 13]). *Let* $A\colon H \to H$ *be a compact self-adjoint operator on a separable Hilbert space* H. *Then* H *can be decomposed as a Hilbert space direct sum*

$$H = H_0 \oplus \bigoplus_{i=1}^{m} H_i,$$

with m *a positive integer or infinity, so that* $A|_{H_i}$ *is equal to multiplication by* μ_i, *for every* $i > 0$ $\dim H_i < \infty$, $\mu_i \neq \mu_j$ *for* $i \neq j$, $\mu_0 = 0$ *and* $\lim_{i\to\infty}\mu_i = 0$ *in case* m *is infinite.*

If $K(x, y)$ is as in Theorem 2.3, Theorem 2.4 implies that the space $L^2(D)$ has an orthonormal basis $\{\phi_i\}_{i=1}^{\infty}$ consisting of eigenfunctions of the integral operator A. We denote by $\{\mu_i\}_{i=1}^{\infty}$ the corresponding eigenvalues. With this notation we have

THEOREM 2.5 (Mercer's theorem, [**H**, Ch. 3, Theorem 17]). *Let* $K(x, y)$ *be as above and assume in addition that the operator* A *defined by* K *on* $L^2(D)$ *is positive* (i.e., $(Af, f) \geqslant 0$ *for every* $f \in L^2(D)$). *Then* $\mu_i \geqslant 0$ *for every* i *and*

$$K(x, y) = \sum_{i=1}^{\infty} \mu_i\phi_i(x)\overline{\phi_i(y)},$$

where the series converges uniformly and absolutely.

Proof of Theorem 2.2. For every $t > 0$ define the operator $P(t)$: $L^2(D) \to L^2(D)$ by the formula

$$P(t)f(x) = \int_D p(x, y, t)f(y)\,dy \tag{2.6}$$

for every $f \in L^2(D)$. It follows from the properties of the Green's function (1.3), (1.5) that Theorem 2.3 is applicable and $P(t)$ is a compact self-adjoint operator for every $t > 0$. (1.3) (c) implies the semi-group property

$$P(t) \circ P(s) = P(t+s). \tag{2.7}$$

Differentiation under the integral shows that, for every $f \in L^2(D)$ the function $v(x, t) = \int_D p(x, y, t)f(y)dy = P(t)f(x)$ satisfies the heat equation in $D \times (0, \infty)$ and vanishes on ∂D. Therefore, by Green's theorem,

$$\begin{aligned} \frac{d}{dt}\|P(t)f\|^2 &= \frac{d}{dt}\int_D |P(t)f(x)|^2 dx \\ &= 2\int_D \Delta P(t)f(x)\overline{P(t)f(x)}dx \\ &= -2\int_D |\nabla P(t)f(x)|^2 dx < 0. \end{aligned} \tag{2.8}$$

Thus, the norm of $P(t)f$ is a nonincreasing function of $t > 0$. If f were continuous and $f \mid \partial D = 0$, then $\lim_{t\to 0^+} P(t)f = f$ uniformly by (1.5) (b). It follows that for such functions

$$\|P(t)f\| < \|f\|. \tag{2.9}$$

Since continuous functions which vanish on ∂D are dense in $L^2(D)$, (2.9) holds for all $f \in L^2(D)$ and it is easily seen that

$$\lim_{t\to 0^+} P(t)f = f \tag{2.10}$$

for every f in $L^2(D)$.

To apply Mercer's theorem we need to know that $P(t)$ is positive for $t > 0$. This follows from the semi-group property (2.7). Indeed

$$(P(t)f, f) = \left(P\left(\frac{t}{2}\right) \circ P\left(\frac{t}{2}\right)f, f\right) = \left(P\left(\frac{t}{2}\right)f, P\left(\frac{t}{2}\right)f\right) \geqslant 0.$$

Now let $\{\phi_i\}_{i=1}^{\infty}$ be the orthonormal basis of $L^2(D)$ consisting of eigenfunctions of $P(1)$, and let $\{\mu_i\}_{i=1}^{\infty}$ be the sequence of corresponding eigenfunctions. Without loss of generality we can assume that all eigenfunctions are real. It follows easily from the semigroup property (2.7) that

$$P\left(\frac{l}{k}\right)\phi_i = \mu_i^{l/k}\phi_i$$

for every positive rational number l/k. By continuity $P(t)\phi_i = \mu_i^t\phi_i$. Observe that $\mu_i > 0$ for every i since

$$0 \neq \phi_i = \lim_{t\to 0^+} P(t)\phi_i = \lim_{t\to 0^+} \mu_i^t\phi_i$$

by (2.10). It follows that $\mu_i = e^{-\lambda_i}$; $\lambda_i > 0$ because of (2.9) and $\lim_{i\to\infty}\lambda_i = \infty$, since $\lim_{i\to\infty}\mu_i = 0$ by Theorem 2.4. By Mercer's theorem

$$p(x, y, t) = \sum_{i=1}^{\infty} e^{-\lambda_i t}\phi_i(x) \cdot \phi_i(y)$$

and the series converges uniformly and absolutely in x and y for every fixed $t > 0$. This implies that convergence is uniform in $x, y \in \bar{D}$ and $t \geq \varepsilon$ for every $\varepsilon > 0$.

To conclude the proof we have to show that $\Delta\phi_i + \lambda_i\phi_i = 0$, and that $\phi_i \mid \partial D = 0$. As remarked above

$$P(t)\phi_i(x) = e^{-\lambda_i t}\phi_i(x)$$

is a solution of the heat equation and vanishes on ∂D. Therefore $\phi_i \mid \partial D = 0$ and

$$-\lambda_i e^{-\lambda_i t}\phi_i(x) = \frac{\partial}{\partial t}\left(e^{-\lambda_i t}\phi_i(x)\right) = \Delta\left(e^{-\lambda_i t}\phi_i(x)\right)$$
$$= e^{-\lambda_i t}\Delta\phi_i(x).$$

After canceling $e^{-\lambda_i t}$, we see that ϕ_i is an eigenfunction of Δ belonging to the eigenvalue λ_i.

3. Distribution of Eigenvalues. We saw above that $\lim_{i\to\infty}\lambda_i = \infty$. Order the eigenvalues so that $\lambda_i \leq \lambda_j$ if $i \leq j$. The celebrated formula of Hermann Weyl states that

$$\lim_{i\to\infty}\frac{\lambda_i^{n/2}}{i} = \frac{C(n)}{\mathrm{vol}(D)}, \tag{3.1}$$

where $C(n) = (4\pi)^{n/2}\Gamma(n/2 + 1)$. To see how the heat equation comes in observe that

$$\sum_{i=1}^{\infty} e^{-\lambda_i t} = \int_D p(x, x, t)\, dx, \tag{3.2}$$

since $\int_D \phi_i(x)^2 dx = 1$, for every i. This follows from Theorem 2.2. The estimates of the Green's function are translated into estimates of $f(t) = \Sigma_{i=1}^{\infty} e^{-\lambda_i t}$, which yield information about eigenvalues. More precisely the rate at which the function $f(t)$ blows up when t approaches zero depends on the rate of growth of eigenvalues.

We begin by establishing an intuitively obvious inequality

$$p(x, y, t) \leq k(x, y, t) \quad \text{for} \quad x, y \in \bar{D}, \quad t > 0, \tag{3.3}$$

and using it to derive a rather crude estimate of the rate of growth of eigenvalues.

$p(x, y, t)$ describes the heat flow in D when the boundary of D is kept at temperature absolute zero. This is the same as the heat flow in the whole space with the exterior of D refrigerated. Clearly the temperature should be less than for the heat flow in the whole space without any cooling, given by $k(x, y, t)$. A rigorous proof is given by recalling that

$$k(x, y, t) - p(x, y, t) = g(x, y, t)$$

satisfies the heat equation in (x, t) and is nonnegative (in fact equal to $k(x, y, t)$ when $x \in \partial D$, and zero for $t = 0$) on the normal boundary $\bar{D} \times \{0\} \cup \partial D \times [0, T]$ for every $T > 0$. The maximum principle implies that $g(x, y, t) \geq 0$, which is the desired inequality. Combining (3.2) and (3.3) we see that

$$\sum_{i=1}^{\infty} e^{-\lambda_i t} = \int_D p(x, x, t)\,dx \leq \int_D k(x, x, t)\,dx$$
$$= \int_D (4\pi t)^{-n/2} dx = (4\pi t)^{-n/2}\mathrm{vol}(D).$$

Let N be a positive integer. Then

$$Ne^{-\lambda_N t} \leq \sum_{i=1}^{N} e^{-\lambda_i t} \leq (4\pi t)^{-n/2}\mathrm{vol}(D).$$

Substitute $t = 1/\lambda_N$ to obtain

$$N \leqslant e(4\pi)^{-n/2}\lambda_N^{n/2}\,\mathrm{vol}(D),$$

i.e.,

$$\lambda_N^{n/2} \geqslant \frac{(4\pi)^{n/2}N}{e\,\mathrm{vol}(D)}.$$

This is not as good as (3.1) but it illustrates the method of proof of (3.1) very nicely with a minimal amount of technicalities. (3.1) will follow from a better estimate of the Green's function in a similar way, except that the translation of the properties of $f(t) = \sum_{i=1}^{\infty} e^{-\lambda_i t}$ near $t = 0$ will be more complicated. The idea behind the proof of (3.1) is the following. For very small values of t, the principle of not feeling the boundary [**K**] tells us that

$$k(x, y, t) \sim p(x, y, t). \tag{3.4}$$

Therefore, by (3.2),

$$\sum_{i=1}^{\infty} e^{-\lambda_i t} \sim \int_D k(x, x, t)\,dx = (4\pi t)^{-n/2}\mathrm{vol}(D),$$

which can be translated into (3.1). To make this rigorous we have to make (3.4) precise. This is done in the following lemma.

LEMMA 3.5 [**W2**]. *For $y \in D$ denote by $l(y)$ the distance from y to the boundary ∂D. For every x and y in D we have*

$$0 \leqslant k(x, y, t) - p(x, y, t) \leqslant \begin{cases} (4\pi t)^{-n/2}\exp\left(-\dfrac{l(y)^2}{4t}\right) & \text{if } 0 < t \leqslant t_0 \\ (4\pi t_0)^{-n/2}\exp\left(-\dfrac{l(y)^2}{4t_0}\right) & \text{if } t \geqslant t_0, \end{cases}$$

where $t_0 = l(y)^2/2n$.

Proof. The first inequality has been already proved. We apply the maximum principle to $g(x, y, t) = k(x, y, t) - p(x, y, t)$ thought of as a function of x and t. The values of $g(x, y, t)$ on the normal boundary of $\bar{D} \times (0, T]$ are given by (1.8) and (1.6). Thus

$$\begin{aligned} g(x, y, t) &\leqslant \sup_{\substack{x \in \partial D \\ 0 < s \leqslant t}} (4\pi s)^{-n/2}\exp\left(-\frac{|x - y|^2}{4s}\right) \\ &\leqslant \sup_{0 < s \leqslant t} (4\pi s)^{-n/2}\exp\left(-\frac{l(y)^2}{4s}\right). \end{aligned}$$

Now the function $(4\pi s)^{-n/2}\exp(-l(y)^2/4s)$ attains its maximum at $s = l(y)^2/2n = t_0$, is decreasing for $s > t_0$, and is increasing for $s < t_0$. This proves the lemma.

Using this lemma we obtain a sharper estimate of $f(t) = \sum e^{-\lambda_i t}$.

THEOREM 3.6. *For every $T > 0$ there exists a constant $c > 0$ such that*

$$0 \leqslant \mathrm{vol}(D)(4\pi t)^{-n/2} - \sum_{i=1}^{\infty} e^{-\lambda_i t} \leqslant Ct^{-n/2+1/2}$$

for $t \in (0, T]$.

Proof. Let D_δ be the set of those $x \in D$ whose distance from ∂D is greater or equal to δ. For

small positive δ, D_δ has a smooth boundary Γ_δ. Fix one such $\delta = \delta_0$.

We saw above that

$$0 \leqslant (4\pi t)^{-n/2}\mathrm{vol}(D) - \sum_{i=1}^{\infty} e^{-\lambda_i t}$$
$$= \int_D (k(x,x,t) - p(x,x,t))\,dx = \int_{D_{\delta_0}} + \int_{A_t} + \int_{B_t},$$

where $A_t = \{x \in D \mid \sqrt{2nt} \leqslant l(x) \leqslant \delta_0\}$ and $B_t = \{x \in \bar{D} \mid 0 \leqslant l(x) \leqslant \sqrt{2nt}\}$ (see Fig. 1). The three integrals above will be estimated separately. The first integral can be estimated, for $0 < t \leqslant \delta_0^2/2n$, using the first inequality in Lemma 3.5.

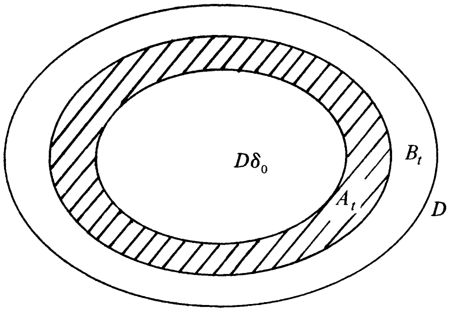


FIG. 1

$$0 \leqslant \int_{D_{\delta_0}} (k(x,x,t) - p(x,x,t))\,dx \leqslant (4\pi t)^{-n/2}\exp(-\delta_0^2/4t)\mathrm{vol}(D), \tag{3.7}$$

which is exponentially small for small t. To estimate the integral over A_t we write it as

$$\int_{\sqrt{2nt}}^{\delta_0}\int_{\Gamma_\delta}(k(x,x,t) - p(x,x,t))\,dS\,d\delta,$$

where dS is the $(n-1)$-dimensional volume element of Γ_δ, and use the first inequality of Lemma 3.5 to estimate the integrand. This yields

$$\int_{A_t}(k(x,x,t) - p(x,x,t))\,dx \leqslant \int_{\sqrt{2nt}}^{\delta_0}\int_{\Gamma_\delta}(4\pi t)^{-n/2}\exp(-\delta^2/4t)\,dS\,d\delta,$$

Clearly the $(n-1)$-dimensional volumes of the boundaries Γ_δ are bounded for $0 \leqslant \delta \leqslant \delta_0$. Therefore

$$\begin{aligned}\int_{A_t}(k(x,x,t) - p(x,x,t))\,dx &\leqslant C \cdot t^{-n/2+1/2}\int_{-\infty}^{\infty}(4\pi t)^{-1/2}\exp\left(-\frac{\delta^2}{4t}\right)d\delta \\ &= C \cdot t^{-n/2+1/2},\end{aligned} \tag{3.8}$$

since $\int_{-\infty}^{\infty}(4\pi t)^{-1/2}\exp(-\delta^2/4t)d\delta = 1$. This estimate holds for $t \in (0, \delta_0^2/2n]$ and the constant C depends only on δ_0. The third integral is estimated in a similar way with the inequality $k(x,x,t) - p(x,x,t) \leqslant k(x,x,t)$ used to estimate the integrand

$$\begin{aligned}\int_{B_t}(k(x,x,t) - p(x,x,t))\,dx &\leqslant \int_0^{\sqrt{2nt}}\int_{\Gamma_\delta}(4\pi t)^{-n/2}\,dS\,d\delta \\ &= (4\pi t)^{-n/2}\int_0^{\sqrt{2nt}}\int_{\Gamma_\delta}dS\,d\delta = C \cdot t^{-n/2+1/2}.\end{aligned} \tag{3.9}$$

As above, the constant C is independent of $t \in (0, \delta_0^2/2n]$. Inequalities (3.7), (3.8), and (3.9) prove the theorem.

Theorem 3.6 implies the formula 3.1 in view of a Tauberian theorem [**Wi**, Theorem 4.3, p. 192] which we state below without proof. Introduce the function $N(\lambda)$, whose value for $\lambda \geqslant 0$ is the number of eigenvalues $\lambda_i \leqslant \lambda$. We can write

$$\sum_n e^{-\lambda_n t} = \int_0^\infty e^{-\lambda_n t} dN(\lambda).$$

THEOREM 3.10. *Suppose $N(\lambda)$ is a nondecreasing function such that $f(t) = \int_0^\infty e^{-\lambda t} dN(\lambda)$ converges for all $t > 0$. If $f(t) \sim A/t^\gamma$ as $t \to 0^+$ for some positive number γ, then*

$$N(\lambda) \sim \frac{A\lambda^\gamma}{\Gamma(\gamma+1)} \quad \text{as } \lambda \to \infty.$$

(Two functions are asymptotic if their ratio approaches one.)

In our case the function $N(\lambda)$ is nondecreasing since the eigenvalues form a nondecreasing sequence and

$$f(t) \sim \operatorname{vol}(D)(4\pi t)^{-n/2}$$

by Theorem 3.6. Therefore

$$N(\lambda) \sim \frac{\operatorname{vol}(D)\lambda^{n/2}}{(4\pi)^{n/2}\Gamma\left(\frac{n}{2}+1\right)}$$

Substitution $\lambda = \lambda_i$ yields (3.1).

References

[**ABP**] M. F. Atiyah, R. Bott, V. K. Patodi, On the heat equation and the index theorem, Invent. Math., 19 (1973) 279-330.

[**BGM**] M. Berger, P. Gauduchon, E. Mazet, Le spectre d'une Variété Riemannienne, Lecture Notes in Math. 194, Springer-Verlag, New York, 1971.

[**F**] A. Friedman, Partial Differential Equations of Parabolic Type, Prentice-Hall, Englewood Cliffs, N.J., 1964.

[**H**] H. Hochstadt, Integral Equations, Wiley, New York, 1973.

[**K**] M. Kac, Can one hear the shape of a drum?, this MONTHLY, 73S (1966) 1-23.

[**L**] E. E. Levi, Sull' equazione del calore, Annali di Mat., 14 (1902) 187-264.

[**LSU**] O. A. Ladyzenskaja, V. A. Solonnikov, N. N. Uralceva, Linear and Quasilinear Equations of Parabolic Type, AMS, Providence, R.I., 1968.

[**McKS**] H. McKean, I. M. Singer, Curvature and the eigenvalues of the Laplacian, J. Differential Geom., 1 (1967) 43-69.

[**M1**] S. Minakshisundaram, A generalization of Epstein zeta function, Canad. J. Math., 1 (1949) 320-327.

[**M2**] S. Minakshisundaram, Eigenfunctions on Riemannian manifolds, J. Indian Math. Soc., 17 (1953) 158-165.

[**MP**] S. Minakshisundaram, A. Pleijel, Some properties of the eigenfunctions of the Laplace operator on Riemannian manifolds, Canad. J. Math., 1 (1949) 242-256.

[**MR**] A. N. Milgram and P. C. Rosenbloom, Harmonic forms and the heat conduction I, Proc. Nat. Acad. Sci. U.S.A., 37 (1951) 180-184.

[**P**] V. K. Patodi, Curvature and the eigenforms of the Laplace operator, J. Differential Geom., 5 (1971) 233-249.

{**PW**] M. N. Protter, H. F. Weinberger, Maximum Principles in Differential Equations, Prentice-Hall, Englewood Cliffs, N. J., 1967.

[**W1**] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte Linearer partieller Differentialgleichungen, Math. Ann., 71 (1911).

[**W2**] H. Weyl, A supplementary note to "A generalization of Epstein zeta function" by S. Minakshisundaram, Canad. J. Math., 1 (1949) 326-327.

[**Wi**] D. V. Widder, The Heat Equation, Academic Press, New York, 1975.

TRANSFINITE SOLUTION TO LAST THEOREM OF FERMAT

PAUL W. DIXON
Department of Psychology, University of Hawaii, Hilo College, Hilo, HI 96720

The last theorem of Fermat asserts that, if n is a natural number greater than 2, then

$$x^n + y^n = z^n$$

has no rational integral solution x, y, z with $xyz \neq 0$.

If this problem is extended so that we consider the transfinite number theory of Cantor (1895), then a solution exists. Thus if each x, y, z is greater than 2, we note

$$x^{\aleph_0} = y^{\aleph_0} = z^{\aleph_0}$$

and using the Cantorian algebra we have

$$x^{\aleph_0} + y^{\aleph_0} = z^{\aleph_0}$$

which is a solution.

This treatment illustrates the asymmetry of the natural numbers, as it is clearly evident that Fermat's problem has no readily apparent solution when enumerating from 1 to $\aleph_0$. But counting down from $\aleph_0$, e.g., $\aleph_0 - 1, \aleph_0 - 2, \aleph_0 - 3, \ldots, \aleph_0 - n$, provides a solution to the problem.

The set of transfinite numbers $\aleph_0 - 1, \aleph_0 - 2, \aleph_0 - 3, \ldots, \aleph_0 - n$ and the positive integers $1, 2, 3, \ldots, n$ are for each corresponding integer $1, 2, 3, \ldots, n$ congruent for the same modulus (namely, 0), thus illustrating the validity of this solution for Fermat's last theorem.

I acknowledge with gratitude the contributions of Arthur A. Sagle to this work.

Reference

1. G. Cantor, Beiträge zur Begrundung der transfinite Mengenlehre I, Math. Ann., 46 (1895), 471–512.

MATHEMATICAL NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

THE TRAPEZOID RULE, STIRLING'S FORMULA, AND EULER'S CONSTANT

RICHARD JOHNSONBAUGH
Department of Mathematics, Chicago State University, Chicago, IL 60628

The purpose of this note is to show how the trapezoid rule can be used to give simple proofs of the existence of the limit in Stirling's formula and the existence of Euler's constant. The method also yields sharp estimates for the remainder.

The trapezoid rule (see, for example, [**10**, p. 218]) states that if f and f' are continuous on $[a, b]$ and f'' exists on (a, b), then

$$\int_a^b f - \frac{(b-a)}{2}\left[f(b) + f(a)\right] = -\frac{f''(\xi)(b-a)^3}{12} \tag{1}$$

for some $\xi \in (a, b)$.

Stirling's formula may be written

$$\lim_{n\to\infty} K_n\sqrt{2\pi} = 1 \tag{2}$$

where $K_n = n^n\sqrt{n}/(n!e^n)$. The crucial part of (2) is the existence of the limit. To prove $\lim_{n\to\infty}K_n$ exists, set $f(t) = \ln t$, $a = k-1$, and $b = k$ in (1) to obtain

$$s_k = \frac{\ln(k-1)+\ln k}{2} - \int_{k-1}^{k} \ln t\, dt = \frac{-1}{12\xi_k^2}$$

where $k-1 < \xi_k < k$. Summing from $k=2$ to $k=n$ gives

$$\sum_2^n s_k = \sum_1^n \ln k - \int_1^n \ln t\, dt - \frac{\ln n}{2} = -1 - \ln K_n. \tag{3}$$

Since $|s_k| < 1/[12(k-1)^2]$, the series Σs_k converges and, thus, $\lim_{n\to\infty}K_n$ exists. Incidentally, it follows from (3) that $\{K_n\}$ is increasing.

References [**2**], [**3**], [**4**], [**5**], [**6**], and [**11**] illustrate a variety of methods of proving that $\lim_{n\to\infty}K_n = 1/\sqrt{2\pi}$.

The remainder, $r_n = -\Sigma_{n+1}^{\infty}s_k$, can be estimated as follows. For $n \geqslant 2$,

$$r_n = \sum_{n+1}^{\infty}\frac{1}{12\xi_k^2} < \sum_{n+1}^{\infty}\frac{1}{12(k-1)^2} < \frac{1}{12}\int_{n-1}^{\infty}\frac{1}{x^2}dx = \frac{1}{12(n-1)}.$$

Similarly,

$$\frac{1}{12(n+1)} < r_n.$$

These estimates were derived in a different way in [**11**, pp. 743–744].

Euler's constant γ is defined as the limit of the sequence $\{\gamma_n\}$ where $\gamma_n = \Sigma_{k=1}^{n}1/k - \ln n$. An argument identical to that above using (1) with $f(t) = 1/t$ gives the existence of $\lim_{n\to\infty}\gamma_n$. The estimate corresponding to that for the remainder in Stirling's formula is

$$\frac{1}{12(n+1)^2} < r_n < \frac{1}{12(n-1)^2}.$$

This implies an estimate obtained by Rao [**9**].

The Euler-Maclaurin formula (see [**1**], [**6**], [**7**], [**8**]), which compares sums and integrals, discloses the relationship among the trapezoid rule, Stirling's formula, and Euler's constant. A simple, special case of the Euler-Maclaurin formula, valid for f and f' continuous on $[a, b]$, is

$$\sum_1^n f(k) - \int_1^n f = \frac{[f(1)+f(n)]}{2} + \int_1^n Pf' \tag{4}$$

where $P(t) = t - [t] - 1/2$. Taking $f(t) = \ln t$ and rearranging terms, (4) may be written

$$\sum_1^n \ln k - n\ln n + n - \frac{\ln n}{2} = 1 + \int_1^n Pf'.$$

Since P is periodic of period 1, $\int_a^{a+1}P = 0$ for all a, and since f' decreases to zero, it follows that $\int_1^{\infty}Pf'$ converges. (The argument is similar to that which shows that $\int_0^{\infty}[(\sin x)/x]dx$ converges.) Therefore,

$$\lim_{n\to\infty}\left[\sum_1^n \ln k - n\ln n + n - \frac{\ln n}{2}\right]$$

exists. The existence of the limit in Stirling's formula has been reestablished. In a similar way, by taking $f(t) = 1/t$, one deduces the existence of γ. Taking $n = 2$ in (4), one obtains

$$\int_1^2 f = \frac{f(1) + f(2)}{2} + R,$$

where $R = -\int_1^2 Pf'$, which is a form of the trapezoid rule.

More elaborate versions of (4) provide powerful, general methods of investigating sums and integrals.

References

1. R. P. Boas, Jr., Partial sums of infinite series, and how they grow, this MONTHLY, 84 (1977) 237–258.
2. R. C. Buck, Advanced Calculus, 3rd ed., McGraw-Hill, New York, 1978.
3. W. Feller, A direct proof of Stirling's formula, this MONTHLY, 74 (1967) 1223–1225.
4. ______, Correction to "A direct proof of Stirling's formula," this MONTHLY, 75 (1968) 518.
5. ______, An Introduction to Probability Theory and Its Applications, vol. 1, 2nd ed., Wiley, New York, 1957.
6. G. H. Hardy, Divergent Series, Oxford University Press, London, 1949.
7. K. Knopp, Theory and Application of Infinite Series, Blackie, London and Glasgow, 1928.
8. D. E. Knuth, The Art of Computer Programming, vol. 1, Addison-Wesley, Reading, Mass., 1968.
9. S. K. L. Rao, On the sequence for Euler's constant, this MONTHLY, 63 (1956) 572–573. (Reprinted in Apostol et al., Selected Papers on Calculus, MAA, 1969, pp. 389–390.)
10. D. F. Riddle, Calculus and Analytic Geometry, 3rd ed., Wadsworth, Belmont, Calif., 1979.
11. A. E. Taylor and W. R. Mann, Advanced Calculus, 2nd ed., Xerox, Lexington, Mass., 1972.

ASYMPTOTIC BEHAVIOR OF NONEXPANSIVE MAPPINGS IN UNIFORMLY CONVEX BANACH SPACES

ELON KOHLBERG

Graduate School of Business Administration, Harvard University, Cambridge, MA 02138

ABRAHAM NEYMAN

Department of Mathematics, University of California, Berkeley, CA 94720

1. Introduction and Statement of the Result. Let X be a uniformly convex Banach space, C a closed convex subset of X, $T\colon C \to C$ a nonexpansive mapping, i.e., $\|Tx - Ty\| \leqslant \|x - y\|$ for all $x, y \in C$.

THEOREM. *$(1/n)T^n x$ converges for all $x \in C$.*

(Note that, since $\|T^n x - T^n y\| \leqslant \|x - y\|$, the limit of $(1/n)T^n x$ must be the same for all $x \in C$.)

Pazy [2] first proved this theorem for Hilbert spaces and raised the question of its validity in uniformly convex spaces. Partial results in this direction were obtained by Reich [3], [4].

Our proof of the theorem is elementary, requiring only a simple geometric argument.

2. Proof of the Theorem. Assume, without loss of generality, that $0 \in C$. Then for every $r > 0$, the contraction mapping $T/(1 + r)$ maps C into C and therefore has a unique fixed point, $x(r)$.

LEMMA 1. *For every $y \in C$ and $r > 0$, $\|rx(r)\| \leqslant 2\|ry\| + \|Ty - y\|$.*

Proof.

$$\begin{aligned}\|x(r) - y\| &\geqslant \|Tx(r) - Ty\| = \|(1 + r)x(r) - Ty\| \\ &\geqslant (1 + r)\|x(r) - y\| - \|ry\| - \|Ty - y\|;\end{aligned}$$

hence $r\|x(r) - y\| \leqslant \|ry\| + \|Ty - y\|$, so that $\|rx(r)\| \leqslant 2\|ry\| + \|Ty - y\|$. Q.E.D.

COROLLARY 2. $\lim_{r \to 0^+} \|rx(r)\|$ *exists and is equal to* $\inf_{y \in C}\|Ty - y\|$.

Proof. By the lemma, $\limsup_{r\to 0^+}\|rx(r)\| \leqslant \inf_{y\in C}\|Ty - y\|$. On the other hand, we have $\inf_{y\in C}\|Ty - y\| \leqslant \inf_{r>0}\|Tx(r) - x(r)\| = \inf_{r>0}\|rx(r)\| \leqslant \liminf_{r\to 0^+}\|rx(r)\|$. Q.E.D.

LEMMA 3. *Let $(X, \|\ \|)$ be a uniformly convex space. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that, if $f \in X^*$ and $x, y \in X$ are such that $\|f\| = 1$, $\|x\|$ and $\|y\|$ are in $[1-\delta, 1+\delta]$, and such that $f(x) \geqslant 1-\delta$ and $f(y) \geqslant 1-\delta$, then $\|x - y\| \leqslant \varepsilon$.*

Proof. By the definition of uniform convexity,

$$\eta \equiv \inf\left\{1 - \left\|\frac{x+y}{2}\right\| : \|x\| = 1, \quad \|y\| = 1, \quad \text{and} \quad \|x-y\| \geqslant \frac{\varepsilon}{3}\right\} > 0.$$

Let $\delta = \min\{\eta/2, \varepsilon/3\}$. By our assumptions,

$$\left\|\frac{1}{2}\left(\frac{x}{\|x\|} + \frac{y}{\|y\|}\right)\right\| \geqslant f\left(\frac{1}{2}\left(\frac{x}{\|x\|} + \frac{y}{\|y\|}\right)\right) \geqslant \frac{1-\delta}{1+\delta} > 1 - \eta.$$

It follows that $\|x/\|x\| - y/\|y\|\,\| \leqslant \varepsilon/3$. But then $\|x - y\| \leqslant \|x/\|x\| - y/\|y\|\,\| + 2\delta \leqslant \varepsilon$. Q.E.D.

The following lemma gives the geometric idea of our proof. It is that, if $\|x\| \gg \|y\|$, if $Tx - x$ is in the same direction as x, and if $\|Ty - y\|$ is not much larger than $\|Tx - x\|$, then $Ty - y$ must be close to $Tx - x$ (for, otherwise, $\|Tx - Ty\| > \|x - y\|$).

LEMMA 4. *For every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\|Ty - y\| \leqslant \inf_{x\in C}\|Tx - x\| + \delta$ then $\|(Ty - y) - rx(r)\| \leqslant \varepsilon$ for all sufficiently small $r > 0$.*

Proof. By Corollary 2, we need consider only the case $\inf_{y\in C}\|Ty - y\| > 0$. Assume, without loss of generality, that $\inf_{y\in C}\|Ty - y\| = 1$.

Let $\delta > 0$ be as in Lemma 3, and let $f_r \in X^*$ be such that $\|f_r\| = 1$ and $f_r(x(r) - y) = \|x(r) - y\|$. If $r > 0$ is sufficiently small, then

$$f_r(rx(r)) = f_r(r(x(r) - y) + ry) = r\|x(r) - y\| + f_r(ry) \geqslant \|rx(r)\| - 2r\|y\| \geqslant 1 - \delta.$$

Since T is nonexpansive, $\|x(r) - y\| \geqslant \|Tx(r) - Ty\| = \|x(r) - y + rx(r) - (Ty - y)\|$; hence $\|x(r) - y\| \geqslant f_r(x(r) - y) + f_r(rx(r)) - f_r(Ty - y)$, which implies $f_r(Ty - y) \geqslant f_r(rx(r))$.

We conclude that $1 + \delta \geqslant f_r(Ty - y) \geqslant f_r(rx(r)) \geqslant 1 - \delta$; by Corollary 2, for a choice of y such that $\|y - Ty\| \leqslant 1 + \delta$, if $r > 0$ is sufficiently small, $1 \leqslant \|rx(r)\| \leqslant 1 + \delta$; since $1 \leqslant \|Ty - y\| \leqslant 1 + \delta$, all the assumptions of Lemma 3 are satisfied, and hence it follows that $\|(Ty - y) - rx(r)\| \leqslant \varepsilon$. Q.E.D.

COROLLARY 5. *$\lim_{r\to 0^+} rx(r)$ exists.*

Proof. If $s > 0$ is sufficiently small, then by Corollary 2, $\|sx(s)\| \leqslant \inf\|Ty - y\| + \delta$. If $r > 0$ is sufficiently small, then, by Lemma 4, $\|sx(s) - rx(r)\| \leqslant \varepsilon$. Thus, $\{rx(r)\}_{r>0}$ is Cauchy. Q.E.D.

Proof of theorem. Denote $a = \inf_{y\in C}\|Ty - y\|$. Given $\varepsilon > 0$, let $\delta > 0$ be as in Lemma 4, and let $y_\varepsilon \in C$ be such that $\|Ty_\varepsilon - y_\varepsilon\| \leqslant a + \delta$; and hence $\|T^{k+1}y_\varepsilon - T^k y_\varepsilon\| \leqslant a + \delta$ for all $k = 0, 1, 2, \ldots$. By Lemma 4, $\|(T^{k+1}y_\varepsilon - T^k y_\varepsilon) - \alpha\| \leqslant \varepsilon$, where $\alpha = \lim_{r\to 0^+} rx(r)$. It follows that $\|(1/n)T^n y_\varepsilon - y_\varepsilon) - \alpha\| \leqslant \varepsilon$ for all $n = 0, 1, 2, \ldots$; hence $\limsup\|(1/n)T^n y_\varepsilon - \alpha\| \leqslant \varepsilon$.

Now, let $x \in X$. Since $\|(1/n)T^n x - (1/n)T^n y_\varepsilon\| \leqslant (1/n)\|x - y_\varepsilon\|$, we have

$$\limsup_{n\to\infty}\|(1/n)T^n x - \alpha\| = \limsup_{n\to\infty}\|(1/n)T^n y_\varepsilon - \alpha\| \leqslant \varepsilon.$$

Since this is true for every $\varepsilon > 0$, it follows that $(1/n)T^n x$ converges to α. Q.E.D.

Without strict convexity, the theorem is false. In fact, we have the following:

PROPOSITION. *Let X be a Banach space and let $f \in X^*$, $\|f\| = 1$, be such that $F = \{x \in X: \|x\| = 1 = f(x)\}$ is a nonempty separable face of the unit ball of X. Then there exists a nonexpansive mapping $T: X \to X$ such that the set of accumulation points of $T^n(0)/n$ equals F.*

Proof. Let $\{z_m\}_{m=1}^{\infty}$ be dense in F and let $\{t_m\}_{m=1}^{\infty}$ be an increasing sequence of positive integers such that $\lim_{m\to\infty} t_{m-1}/t_m = 0$. Define $\gamma: R \to X$ by

$$\gamma(t) = \begin{cases} tz_1 & \text{if } t \leq t_1, \\ \gamma(t_{m-1}) + (t - t_{m-1})z_m & \text{if } t_{m-1} < t \leq t_m, \end{cases}$$

and $T: X \to X$ by $Tx = \gamma(f(x) + 1)$. The mapping T is nonexpansive, for $\|z_m\| = 1$ for all $m = 1, 2, \ldots$, and therefore $\|\gamma(t) - \gamma(s)\| \leq |t - s|$ for all $t, s \in R$. Hence $\|Tx - Ty\| = \|\gamma(f(s) + 1) - \gamma(f(y) + 1)\| \leq |f(x) - f(y)| \leq \|x - y\|$.

Note that $f(z_m) = 1$ for all $m = 1, 2, \ldots$, and therefore $f(\gamma(t)) = t$. It follows that $T^n(0) = \gamma(n)$ so that $T^n(0)/n = \gamma(n)/n$ is a convex combination of points in F, and all the accumulation points of $\{T^n(0)/n\}_{n=1}^{\infty}$ lie in F. On the other hand,

$$\left\| \frac{\gamma(t_m)}{t_m} - z_m \right\| = \left\| \frac{\gamma(t_{m-1})}{t_m} + \frac{t_m - t_{m-1}}{t_m} z_m - z_m \right\| \to 0,$$

which implies that every point in F is an accumulation point of $\{\gamma(t_m)/t_m\}_{m=1}^{\infty}$ and hence of $\{T^n(0)/n\}_{n=1}^{\infty}$. Q.E.D.

We have seen that uniform convexity is a sufficient condition, whereas strict convexity is a necessary condition for the convergence of $T^n(0)/n$. It turns out ([**1**]) that a necessary and sufficient condition is the Fréchet differentiability of the norm of X^*.

The work for this paper was supported by National Science Foundation Grant MCS-79-06634.

References

1. E. Kohlberg and A. Neyman, Asymptotic behavior of non-expansive mappings in normed linear spaces, Israel J. Math. (to appear).

2. A. Pazy, Asymptotic behavior of contractions in Hilbert spaces, Israel J. Math., 9 (1971) 235–240.

3. S. Reich, Asymptotic behavior of contractions in Banach spaces, J. Math. Anal. Appl., 44 (1973) 57–70.

4. S. Reich, Asymptotic behavior of semigroups of nonlinear contractions in Banach spaces, J. Math. Anal. Appl. 53 (1976), 277–290.

CLASSROOM NOTES

EDITED BY DEBORAH TEPPER HAIMO AND FRANKLIN TEPPER HAIMO

Material for this department should be sent to Professor Deborah Tepper Haimo, Department of Mathematical Sciences, University of Missouri, St. Louis, MO 63121.

CAUCHY SEQUENCES AND FUNCTION RINGS

STEPHAN C. CARLSON
Department of Mathematics, Grinnell College, Grinnell, IA 50112

Let C denote the ring of continuous, real-valued functions defined on the closed unit interval

$[0,1]$, and for $p \in [0,1]$ let A_p denote the ideal of C consisting of those functions in C that vanish at p. The fact that the maximal ideals of C are precisely the ideals of the form A_p for some $p \in [0,1]$ provides an interesting example in a first undergraduate-level course in abstract algebra [**2**, p. 139] and a glimpse at one of the basic relationships between algebra and analysis. The feature of $[0,1]$ that is essential in the proof of the more difficult implication of this characterization (that every maximal ideal is of the form A_p) is its compactness [**1**, p. 58]. However, the compactness of $[0,1]$ may be exploited in an equivalent form: the usual metric on $[0,1]$ is complete and totally bounded. Applying the concepts of Cauchy sequences and completeness, which are of fundamental importance in calculus, to characterize the maximal ideals of C seems quite appropriate at the undergraduate level.

The purpose of this note is to present a proof that (1) treats the completeness of $[0,1]$ as its essential ingredient and (2) emphasizes the interplay between the algebraic properties of C and the behavior of the continuous functions on $[0,1]$.

THEOREM. *If A is a maximal ideal of C, then there is a point $p \in [0,1]$ such that $A = A_p$.*

Proof. Note for reference that each member of any proper ideal of C must vanish at some point of $[0,1]$. (A member of C that never vanishes on $[0,1]$ is an invertible member of C.) Also note that if the sum of two nonnegative functions vanishes at some point then each of the two functions vanishes at that point.

Let n be a positive integer and for $1 \leq i \leq n$ let h_i be a nonnegative member of C such that $h_i(x) = 0$ if and only if $(i-1)/n \leq x \leq i/n$. Since C is a commutative ring with identity, every maximal ideal of C is a prime ideal [**2**, p. 167]. So A is a prime ideal, and, since $h_1 \cdot h_2 \cdots h_n = 0$, $h_i \in A$ for some i. Rename this h_i as f_n.

In this way for each positive integer n we can construct a nonnegative function $f_n \in A$ whose zero set is a closed interval $[a_n, b_n] \subseteq [0,1]$ of length $1/n$ (i.e., $f_n(x) = 0$ if and only if $x \in [a_n, b_n]$). Now if m and n are two positive integers then $f_m + f_n \in A$. So there is a point $x_{m,n} \in [0,1]$ for which $f_m(x_{m,n}) + f_n(x_{m,n}) = 0$, and hence $x_{m,n} \in [a_m, b_m] \cap [a_n, b_n]$. Using these facts and the completeness of $[0,1]$, we can easily show that the sequences $\{a_n\}$ and $\{b_n\}$ are Cauchy and have a common limit $p \in [0,1]$.

Now suppose $f \in A$. For each positive integer n, $f^2 + f_n \in A$. So there is a point $x_n \in [0,1]$ for which $f(x_n)^2 + f_n(x_n) = 0$, and hence $x_n \in [a_n, b_n]$ and $f_n(x_n) = 0$. Since $x_n \to p$, $f(p) = 0$ by the continuity of f. Thus, $f \in A_p$.

Therefore, $A \subseteq A_p$, and, since A is a maximal ideal, $A = A_p$.

Note. The total boundedness of $[0,1]$ is used in the proof in a completely transparent fashion. Let X be any complete and totally bounded metric space, let $C(X)$ be the ring of continuous real-valued functions on X, and let A be a maximal ideal of $C(X)$. Then we can use the total boundedness of X to construct nonnegative functions $f_n \in A$ with zero sets Z_n of diameter less than or equal to $1/n$. It can be shown that there is a point $p \in X$ which is the common limit of all sequences whose nth term lies in Z_n. The remainder of the proof translates directly to obtain $A = A_p$.

References

1. L. Gillman and M. Jerison, Rings of Continuous Functions, D. Van Nostrand, Princeton, N.J., 1960.
2. I. N. Herstein, Topics in Algebra, 2nd ed., Wiley, New York, 1975.

THE PLANIMETER AS AN EXAMPLE OF GREEN'S THEOREM

RONALD W. GATTERDAM
Department of Mathematics, University of Wisconsin–Parkside, Kenosha, WI 53141

The purpose of this note is to describe the use of Green's Theorem to explain the working of the polar (or Amsler's) planimeter in computing the area of a plane figure. While there are other,

and so is equal to the area of M. It should be observed that the computation above demonstrates the fact that, in computing the line integral over a closed path, C,

$$\oint_C (F+G) = \oint_C F$$

for any conservative function G.

As a final remark to the instructor wishing to use this material in the classroom, obtain a planimeter and demonstrate or allow the students to use it to see what a simple mechanical device it is. To locate a planimeter, try your geography or civil engineering departments.

References

1. R. Courant, Differential and Integral Calculus, vol. 2, Interscience, New York, 1936.
2. Compensating Polar Planimeter Instruction Manual, Keuffel and Esser Co., New York, 1963.

PROBLEMS AND SOLUTIONS

EDITED BY J. L. BRENNER, VLADIMIR DROBOT, AND ROGER C. LYNDON

Send all ***proposed*** *problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.

Solutions *should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

SOLUTIONS OF PROBLEMS DEDICATED TO EMORY P. STARKE

Transformed Euler Differential Equation

S 28 [1980, 218]. *Proposed by David A. Sánchez, University of New Mexico.*

Find the general solution of the differential equation

$$x^3y'' + 2x^2 = (xy' - y)^2.$$

Solution by Deborah Frank Lockhart, Michigan Technological University. Let $y = -x\ln|v|$. Then $x^2v'' + 2xv' - 2v = 0$, which is an Euler equation with solution $v = \alpha x^{-2} + \beta x$. Hence

$$y = -x\ln|\alpha x^{-2} + \beta x|.$$

Note. Several solvers noted that absolute value signs can be dropped if the variables are complex. Some kept track of one parameter families of "singular" solutions "lost" in certain manipulations, but E. V. Norrset noted using the Ritt Theory of algebraic differential equations

that all solutions are contained in the general solution.

Also solved by 53 other readers.

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by March 31, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2908. *Proposed by Ken Brown, Nova High School, Ft. Lauderdale, Fla.*

Let $P(x) = 2x^3 + 3x^2 - 1$. Define the sequence $\{a_n\}$ by

$$a_1 = 4, a_{n+1} = P(a_n) - 10^f[P(a_n)/10^f], \qquad (n > 0, f = 2^n).$$

Here [ ] denotes the greatest integer function. Show that each a_n satisfies the congruence $x^3 \equiv x \bmod 10^e$, $e = 2^{n-1}$.

E 2909. *Proposed by I. M. Isaacs and Mark Manasse, University of Wisconsin.*

Let us say that a set of nonnegative real numbers is "closed under $\pm$" if for every $x, y \in S$, either $x + y \in S$ or $|x - y| \in S$. For instance, if $\alpha > 0$ and $n \geqslant 0$ is an integer, then the set $S(n, \alpha) = \{0, \alpha, 2\alpha, \dots, n\alpha\}$ has this property. Show that, if a finite set is closed under $\pm$ and it is not of the form $S(n, \alpha)$, then the set contains exactly four elements.

E 2910. *Proposed by Tian Jing Huang, Szechwan University, Peoples Republic of China*

Prove that the function $A(h) \equiv \int_0^k[1 - (h - \cos x)^2]^{1/2}[4(h - \cos x)^2 - 1]\,dx$, $k = \arccos(h - 1)$, has at least two zeros in the interval $(0, 2)$.

E 2911. *Proposed by Jordi Dou, Barcelona, Spain.*

Let 2 semicircles AC, CB, $\overline{AC} = 3\overline{CB}$ be given (ACB are collinear). Let a, b be the tangents to the given semicircles at A, B. Let γ be the circle tangent to a, b and to the larger of the given semicircles. Prove that γ, b, and the given semicircles have a common tangent circle.

E 2912. *Proposed by Barry Powell, Kirkland, Washington.*

Let N_{2m}, D_{2m} be the numerator and denominator of the $2m$th Bernoulli number B_{2m}, defined by $\tan z = \Sigma(-1)^{n-1}2^{2n}(2^{2n-1} - 1)B_{2n}z^{2n-1}/(2n)!$.

Prove: (a) If p is an odd prime, $N_{2p}/p \equiv 1 \bmod p$. (b)* For any even positive integer k, there exist infinitely many even positive integers $r, s, t, \dots$ such that $N_k \mid N_r$, $N_k \mid N_s, \dots$.

E 2913. *Proposed by John J. Wahl, Mt. Pocono, Pa.*

Given rational N and P, find all integers a, b, x, y satisfying $a + y = b + x$, for which $N = (ay - bx)/(xy - ab)$, $P = (xy - ab)/(ax - by)$.

SOLUTIONS OF ELEMENTARY PROBLEMS

A Divergent Partial Sum

E 2744 [1978, 823; 1979, 503]. *Proposed by H. L. Montgomery, University of Michigan.*

Let $a_n \geqslant 0$ and $a_{m+n} \leqslant a_m + a_n$ for $m, n = 1, 2, \dots$. Show that $\Sigma_{k=1}^n k^{-2}a_k \geqslant \frac{1}{4}n^{-1}a_n \log n$.

Solution by Gustaf Gripenberg, Helsinki University of Technology, Finland. The following

sharper result holds: Let $a_1, a_2, a_3, \ldots$ be a sequence of real numbers satisfying $a_{m+n} \leqslant a_m + a_n$, $m, n \geqslant 1$. Then

$$\sum_{k=1}^{n} k^{-2}a_k \geqslant n^{-1}a_n \Sigma k^{-1} > n^{-1}a_n \log n. \tag{1}$$

For any $m \geqslant 2$, the assumption implies $a_m \leqslant a_k + a_{m-k}$ for $k = 1, 2, \ldots, m-1$. Adding these inequalities and solving for a_m yields

$$a_m \leqslant 2(m-1)^{-1} \sum_{i=1}^{m-1} a_i, \qquad m \geqslant 2. \tag{2}$$

Define numbers y_m by

$$y_m = 2m^{-1}(m-1)^{-1} \sum_{1}^{n-1} k^{-1}, \qquad 2 \leqslant m \leqslant n; \tag{3}$$

define $y_{n+1} = 0$. Let

$$x_m = y_m - y_{m+1}, \qquad 2 \leqslant m \leqslant n; \tag{4}$$

define $x_1 = 0$. Then

$$\sum_{m+1}^{n} x_i = y_{n+1}, \qquad 1 \leqslant m \leqslant n-1. \tag{5}$$

If (3) is substituted into (4), x_m can be expressed as $x_m = 2(m-1)^{-1}(y_{m+1} - m^{-2})$, where $2 \leqslant m \leqslant n-1$. This, together with (5), yields

$$(m-1)x_m = 2\left(\sum_{m+1}^{n} x_i - m^{-2}\right), \qquad 1 \leqslant m \leqslant n-1. \tag{6}$$

It follows by backwards induction on $m = n-1,\ n-2, \ldots, 1$, that

$$n^{-1}a_n \sum_{1}^{n} k^{-1} - \sum_{1}^{n} k^{-2}a_k \leqslant \sum_{1}^{m}\left(\sum_{m+1}^{n} x_i - k^{-2}\right) a_k, \tag{7}$$

the (outer) sums being on k. Indeed, (2) with $m = n$ and the definition of x_n by (3) and (4) imply

$$n^{-1}a_n\left(\sum_{1}^{n} k^{-1} - n^{-1}\right) \leqslant n^{-1}(n-1)^{-1}2\left(\sum_{1}^{n-1} a_k\right)\left(\sum_{1}^{n-1} k^{-1}\right) = x_n \sum_{1}^{n-1} a_k.$$

Subtracting $\Sigma_1^{n-1} k^{-2}a_k$ from both sides yields (7) in the case $m = n-1$. Assuming (7) is true for m $(n-1 \geqslant m \geqslant 2)$, (6) and (2) imply

$$\begin{aligned}
\sum_{1}^{n}\left(\sum_{m+1}^{n} x_i - k^{-2}\right) a_k &= \sum_{1}^{m-1}\left(\sum_{m+1}^{n} x_i - k^{-2}\right) a_k + (m-1)x_m a_m/2 \\
&\leqslant \sum_{1}^{m-1}\left(\sum_{m+1}^{n} x_i - k^{-2}\right) a_k + \sum_{1}^{m-1} x_m a_k \\
&= \sum_{1}^{m-1}\left(\sum_{m}^{n} x_i - k^{-2}a_k\right).
\end{aligned}$$

This completes the induction.

If $m = 1$ is substituted into (7), the right-hand side becomes

$$\left(\sum_{i=2}^{n} x_i - 1\right) a_1 = (y_2 - 1)a_1 = 0.$$

The assertion and claim follow.

Also solved by D. Hensley, L. E. Mattics, James Theiler, and the proposer.

Inversion of the Incenter, Circumcenter, Nine-points Center

E 2793 [1979, 703]. *Proposed by E. D. Camier, Merseyside, England.*

P and Q are two points isogonally conjugate with respect to a triangle ABC of which the circumcenter, orthocenter, and nine-points center are O, H, and N, respectively. If $\overrightarrow{OR} = \overrightarrow{OP} + \overrightarrow{OQ}$, and U is the point symmetric to R with respect to N, show that the isogonal conjugate of U in the triangle ABC is the intersection V of the lines P_1Q and PQ_1 where P_1 and Q_1 are the inverses of P and Q in the circle ABC. (Assume that neither of P, Q is on the circle ABC.)

Solution by Sister Stephanie Sloyan, Georgian Court College, Lakewood, N.J. I. Let the triangle be inscribed in a unit circle and use complex coordinates with $s_1 = a + b + c$, $s_2 = ab + ac + bc$, and $s_3 = abc$. Note $s_2 = \bar{s}_1 s_3$. Because P and Q are two points isogonally conjugate with respect to triangle ABC, $p + q + s_3\bar{p}\bar{q} = s_1$. Since $\overrightarrow{OR} = \overrightarrow{OP} + \overrightarrow{OQ}$, $r = p + q$. H, the orthocenter, has coordinate s_1; and N, the nine-points center, has coordinate $s_1/2$. If U is the point symmetric to R with respect to N, then $\frac{1}{2}[(p + q) + u] = \frac{1}{2}s_1$, or $u = s_1 - (p + q)$. If V_1 is the isogonal conjugate of U in the triangle ABC, then

$$v_1 + u + s_3\bar{v}_1\bar{u} = s_1. \qquad (*)$$

Equating the two values of $\bar{v}_1$ obtained from $(*)$ and from its conjugate, it is straightforward to obtain the relation

$$v_1(1 - u\bar{u}) = s_1 - u - s_2\bar{u} + s_3\bar{u}^2 = s_1 - u - s_3\bar{u}\left[(s_2 - s_3\bar{u})/s_3\right].$$

Then, using $s_3\bar{s}_3 = 1$, the two equations $u = s_1 - p - q = s_3\bar{p}\bar{q}$, and their two conjugates $s_3\bar{u} = s_2 - s_3(\bar{p} + \bar{q}) = pq$, lead to

$$v_1 = [p + q - pq(\bar{p} + \bar{q})]/(1 - p\bar{p}q\bar{q}). \qquad (**)$$

II. Since P_1 is the inverse of P in the circle ABC, its coordinate is $1/\bar{p}$ and the line P_1Q is represented by

$$z\bar{p}(1 - p\bar{q}) - \bar{z}p(1 - q\bar{p}) + p\bar{q} - \bar{p}q = 0. \qquad (\dagger)$$

Similarly, Q_1 has coordinate $1/\bar{q}$ and Q_1P is given by

$$z\bar{q}(1 - \bar{p}q) - \bar{z}q(1 - p\bar{q}) + q\bar{p} - \bar{q}p = 0. \qquad (\ddagger)$$

Now solve $(\dagger)$ and $(\ddagger)$ simultaneously to get $v_2 = [p + q - pq(\bar{p} + \bar{q})]/(1 - p\bar{p}q\bar{q})$, so that $v_2 = v_1$.

Also solved by Aage Bondesen (Denmark), C. F. Parry (England) (with a synthetic proof), and the proposer.

The proposer mentions a special case. If P, Q are the foci of the conic tangent to BC, CA, AB at their midpoints, then U and V are respectively the centroid G and the symmedian point K of triangle ABC. The line P_1Q_1 cuts circle ABC in two points: F, the midpoint of P_1Q_1, and J, on the line OK. F is the Feuerbach point (point of contact of the nine points circle and the incircle) of the triangle formed by the tangents to circle ABC at the vertices A, B, C.

Inversion in the circle ABC shows that $PQOF$ are concyclic and form a harmonic quadrangle, so that PQ bisects angle OGF.

The Simson lines of J and F are respectively perpendicular to PQ, and parallel to the Euler line OG of triangle ABC.

Geometric Application of the Inequality $\cos\theta + (1 - \cos\theta)/n > \cos((1 - 1/(2n))\theta)$

E 2816 [1980, 136]. *Proposed by R. Bojanic, Ohio State University.*

Consider a circular segment AOB with $\measuredangle AOB < \pi$. Let C be the orthogonal projection of the point B on the line $\overline{OA}$. Suppose that the arc $\widehat{AB}$ and the segment $\overline{CA}$ are each divided into n

equal parts. If M is the point of the partition of the arc AB closest to B, and N the point of the partition of the segment $\overline{CA}$ closest to C, show that the projection of the midpoint of the arc $\widehat{MB}$ onto the line $\overline{OA}$ is always contained in the interval (C, N).

I. *Solution by Noel Glick, student, Brooklyn College.* Let $\measuredangle AOB$ be θ. It is easily seen that the assertion is equivalent to the inequality

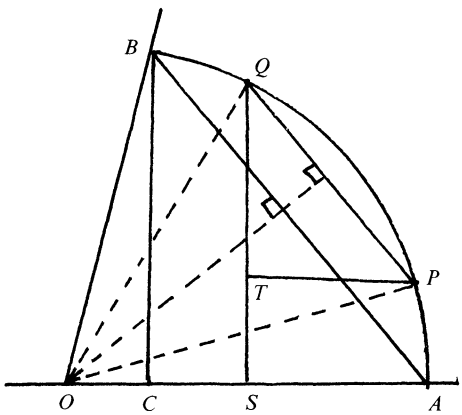
$$\cos\theta + \frac{1}{n}(1 - \cos\theta) > \cos\left(1 - \frac{1}{2n}\right)\theta.$$

More generally, we show that $f(\theta) = \cos\theta + 2a(1 - \cos\theta) - \cos(1 - a)\theta$ is positive on $(0, \pi)$ for any a in $(0, 1)$. Clearly $f(0) = 0$; so it suffices to show $f'(\theta) > 0$ on $(0, \pi)$. Now,

$$f'(\theta) = (2a - 1)\sin\theta + (1 - a)\sin(1 - a)\theta.$$

From the downward concavity of the sine function on $(0, \pi)$, $\sin(1 - a)\theta > (1 - a)\sin\theta$. Hence $f'(\theta) > [(2a - 1) + (1 - a)^2]\sin\theta = a^2\sin\theta > 0$.

II. *Solution by Dalton E. Orr, University of South Alabama.* Let Q be the midpoint of arc MB; let P be the point in arc AB for which arc AP and arc QB have the same length. Note that $n \cdot \text{arc } PQ = (n - 1)\,\text{arc}AB$. (See figure.)



Let S be the orthogonal projection of Q onto OA, and T the orthogonal projection of P onto QS. Let $\alpha = |\measuredangle POQ|$ and $\beta = |\measuredangle AOB|$.

$$\frac{|PQ|}{|AB|} = \frac{2|AO|\sin\alpha/2}{2|AO|\sin\beta/2} = \frac{\sin\alpha/2}{\sin\beta/2} \geqslant \frac{\alpha}{\beta} = \frac{n-1}{n}$$

The inequality is seen by considering $f(x) = \sin x/\sin\gamma - x/\gamma$ for $0 < \gamma < \pi/2$ and $0 \leqslant x \leqslant \gamma$.

Since ΔPTQ and ΔACB are similar, and $|AS| > |PT|$, it follows that

$$\frac{|AS|}{|AC|} > \frac{|PT|}{|AC|} = \frac{|PQ|}{|AB|} \geqslant \frac{n-1}{n}.$$

Also solved by A. Venetoulias & A. Matsoukas (Greece), and the proposer.

Average Velocity

E 2830 [1980, 404]. *Proposed by R. P. Boas, Northwestern University.*

Let a point move on a straight line with steadily increasing acceleration from time $t = 0$ to $t = T$. Show that its velocity at mid-time ($t = T/2$) cannot exceed its average velocity (meaning, as usual, total distance divided by total time). This statement is not necessarily true if $1/2$ is replaced by any other number between 0 and 1.

I. *Solution by Kenneth A. Klinger, Chicago, Ill.* Since the acceleration, $a(t)$ say, is steadily increasing, it follows that any antiderivative of it is convex. If $v(t)$ is the velocity then $v(t) = v(0) + \int_0^t a(x)\,dx$, $0 \leqslant t \leqslant T$. Applying Jensen's inequality to $v(t)$ gives $v(T/2) \leqslant 1/T\int_0^T v(x)\,dx$.

II. *Solution by Mark D. Meyerson, U.S. Naval Academy.* If $v(t)$ is the velocity, let $f(t) = v(t) + v(T-t)$, the sum of v and its reflection in $t = T/2$. Then $f'(t) = v'(t) - v'(T-t)$, which is increasing; moreover $f'(T/2) = 0$. Hence f has minimum at $T/2$; so $v(T/2) = f(T/2)/2 \leqslant$ (average of $f(t))/2$ = average velocity. For a counterexample, when $t \neq T/2$ consider $v(t) = \pm t + \varepsilon t^2$ with $\varepsilon > 0$ sufficiently small.

Also solved by U. Abel (West Germany), E. Adams, K. F. Andersen, A. Berman & M. Pachter (South Africa), I. Bivens, D. M. Bloom, C. E. Bredlau, R. Breusch, R. J. Bridgman, F. S. Cater, Chico Problem Group, J. Cruthirds & J. Morrow, R. Cuculiere (France), T. E. Elsner, L. Erlebach, A. Fieldsteel, G. Fisher, D. M. Friedlen, N. Glick, A. Guzman, E. Heil & J. Schaaf (West Germany), V. Hernandez (Spain), E. D. Huthnance, E. Johnston, M. Josephy, M. S. Klamkin, M. F. Kruelle (student), L. Kuipers (Switzerland), G. Letac (France), G. N. Lewis, J. T. Lewis, O. P. Lossers (Netherlands), R. Lyons, R. Megginson, D. K. Mick, W. Myers, D. E. Orr, M. R. Railkar (India), O. G. Ruehr, M. B. Ruskai, St. Olaf Problem Group, J. Schaer (Canada), R. Sheets, P. S. Shoenfeld, M. Skalsky, M. Solc (Czechoslavakia), A. H. Stein, G. R. Taylor, James W. Thomas & Susan Yeh, W. R. Utz, J. Wiener, H. Ziehms (West Germany), and the proposer.

Diagonals of a Hexagon

E 2831 [1980, 404]. *Proposed by M. Cavachi, University of Bucharest, Rumania.*

Prove that a convex hexagon with no side longer than 1 unit must have at least one main diagonal not longer than 2 units.

I. *Solution by S. A. Belbas, University of Maryland.* We use the well-known fact that for a convex quadrilateral, the product of the diagonals does not exceed the sum of the products of opposite sides.

Let the hexagon be lettered $ABCDEF$ in such a way that $AE \leqslant AC \leqslant CE$. Applying the theorem above to quadrilateral $AEDC$, we have

$$AD \cdot CE \leqslant AE \cdot CD + AC \cdot DE \leqslant AE + AC \leqslant 2CE,$$

from which $AD \leqslant 2$.

II. *Solution by L. Kuipers, Switzerland.* Applying the theorem cited above,

$$\begin{aligned}
BE \cdot AC &\leqslant BC \cdot AE + AB \cdot EC \leqslant AE + EC\\
BE \cdot FD &\leqslant DE \cdot BF + EF \cdot BD \leqslant BF + BD\\
AD \cdot FB &\leqslant AB \cdot FD + AF \cdot BD \leqslant FD + BD\\
AD \cdot EC &\leqslant CD \cdot AE + DE \cdot AC \leqslant AE + AC\\
CF \cdot BD &\leqslant BC \cdot FD + CD \cdot FB \leqslant FD + FB\\
CF \cdot AE &\leqslant EF \cdot AC + FA \cdot EC \leqslant AC + EC.
\end{aligned}$$

Summing and rearranging,

$$(FD + AC)(2 - BE) + (FB + EC)(2 - AD) + (BD + AE)(2 - CF) \geqslant 0.$$

At least one of the three terms must be positive; hence BE or AD or $CF \leqslant 2$.

III. *Solution by Jordi Dou, Barcelona, Spain.* Among the vectors $\overrightarrow{AB}, \overrightarrow{CD}$, and $\overrightarrow{EF}$, some two meet at angle $\leqslant 120°$. Let them be $\overrightarrow{AB}$ and $\overrightarrow{CD}$. Choose point M such that $\overrightarrow{BM} = \overrightarrow{CD}$. Then $MD = BC \leqslant 1$. Since $\measuredangle ABM \leqslant 60°$, and $AB \leqslant 1$, $MB \leqslant 1$, we have by the law of cosines that $AM \leqslant 1$. Thus $AD \leqslant AM + MD \leqslant 2$.

Also solved by R. Breusch, G. Gagola, L. Gesing (Hong Kong), C. Jantzen, M. S. Klamkin (Canada), F. Krieger (Austria), O. P. Lossers (Netherlands), N. Martin (student, Canada), J. Schaer, M. Wolterman, an anonymous solver, and the proposer.

Polynomial Algorithm for Sets

E 2838 [1980, 489]. *Proposed by U. Abel and F. Boukal, University of Bielefeld, West Germany.*

Let X be a measure space with measure m. For any family $\mathcal{A}$ of measurable subsets of X define $m(\mathcal{A}) = m(\cup_{A \in \mathcal{A}} A)$. Let $\mathcal{C}$ be a covering of X by measurable subsets and let k be the minimal cardinality of a subcollection of $\mathcal{C}$ that covers X. Prove that for any $c > 0$ there exists a polynomial algorithm for finding a subcollection $\mathcal{B}$ of $\mathcal{C}$ of cardinality k so that $m(\mathcal{B}) \geqslant (1 - c^{-1})m(X)$.

Solution by the proposers. From the hypothesis it is clear that for any measurable $A \subset X$ there is a $C \in \mathcal{C}$ with $m(C \cap A) \geqslant 1/km(A)$. Thus it is possible to choose $B_1, \dots, B_k$ from $\mathcal{C}$ successively such that

$$m\left(B_i \cap \left(X \setminus \bigcup_{j<i} B_j\right)\right) \geqslant \frac{1}{k} m\left(X \setminus \bigcup_{j<i} B_j\right).$$

Put $B = \{B_1, \dots, B_k\}$ and let r be the following sum of k terms.

$$r = 1/k + (1 - 1/k)1/k + (1 - 1/k - (1 - 1/k)1/k)1/k + \cdots.$$

Then obviously $m(B) \geqslant m(X) \cdot r$. If $a = 1/k$, then by induction it is easy to prove that

$$r = a + (1 - a)a\left[1 + (1 - a) + \cdots + (1 - a)^{k-2}\right].$$

This implies that

$$r = a + (1 - a)a\frac{(1 - a)^{k-1} - 1}{-a} = 1 - (1 - 1/k)^k \geqslant 1 - 1/e,$$

from which the proposition follows.

Also solved by F. S. Cater.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor Roger C. Lyndon, Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (USA), by March 31, 1982. The solver's full post-office address should be on each sheet.

6362. *Proposed by F. S. Cater, Portland State University.*

Let $F(z)$ be an analytic function for $|z| < 1$. For each number u, $0 \leqslant u < 2\pi$, let

$$h(u) = \sup_{0<t<1} |F(te^{iu})| \quad \text{and} \quad g(u) = \lim_{t \to 1-} \sup |F(te^{iu})|.$$

Suppose that for each positive integer n, $\int_0^{2\pi} h(u)^n du < \infty$ in the Lebesgue sense. (This would happen, for example, if F were bounded.) Prove that

$$\sup_{0 \leqslant u < 2\pi} g(u) = \operatorname*{essential\,sup}_{0 \leqslant u < 2\pi} g(u) = \sup_{|z|<1} |F(z)|.$$

6363. *Proposed by H. G. Diamond, University of Illinois, and P. Erdős, Hungarian Academy of Sciences.*

Let ϕ denote Euler's function and set

$$D(\alpha) = \lim_{x \to \infty} \frac{1}{x} \#\{n \leqslant x : \phi(n)/n \leqslant \alpha\},$$

the distribution function of $\phi(n)/n$. Prove that

$$D\left(\frac{1}{2}\right) - D\left(\frac{1}{4}\right) + D\left(\frac{1}{8}\right) - D\left(\frac{1}{16}\right) + \cdots = \frac{1}{2}.$$

6364. *Proposed by the late K. B. Leisenring, University of Michigan.*

A circle with center at the vertex and radius equal to the latus rectum meets a parabola at P, Q. The circle and parabola have common tangents meeting the parabola at X, Y. Prove that XP, YQ are tangent to the circle.

6365. *Proposed by Alan Wilde, University of Michigan.*

For integers $n > h \geqslant 0$, define $\exp_h(z) = \Sigma z^k/k!$, where the sum is over all nonnegative integers $k \equiv h \pmod{n}$.

(i) Assuming Fermat's Last Theorem, show that if $n > 2$ there is no complex number z such that $\exp_0(z)$ and $\exp_1(z)$ are nonzero rational numbers while $\exp_2(z) = \cdots = \exp_{n-1}(z) = 0$.

(ii)* Is Fermat's Last Theorem needed?

6366. *Proposed by Emeric Deutsch, Polytechnic Institute of New York,*

Let A be a row stochastic matrix such that $\|A\| = 1$, $\|\ \|$ being the operator norm induced by the Euclidean vector norm. Show that A is doubly stochastic.

6367. *Proposed by A. Ehrenfeucht and J. Mycielski, University of Colorado.*

Let A be a finite collection of distinct but possibly overlapping regular n-gons of the same size on the plane such that every vertex of every n-gon of A is a vertex of exactly two n-gons of A.

(a) Construct a collection A of $2n$ n-gons such that, even if the n-gons are rigid, A is flexible.

(b)* For which n is a rigid A possible?

SOLUTIONS OF ADVANCED PROBLEMS

Functional Equation $f(f(x)) = a + bx$

6287 [1980, 65]. *Proposed by Brook Taylor, Phoenix, Arizona.*

What is the smallest integer n with the following property? There exist a partition of $\mathbf{R}$ into n sets D_i and n real analytic functions f_i, each f_i defined on some open set containing D_i, such that, if we define a function f: $\mathbf{R} \to \mathbf{R}$ by setting

$$f(x) = f_i(x) \qquad \text{for } x \text{ in } D_i,$$

then $f(f(x)) = -x$ for all x in $\mathbf{R}$.

I. *Solution by Brian R. Hunt, sophomore, Montgomery Blair High School, Silver Spring, Maryland.* Let $D_1 = \{x\colon [-|x|] \text{ is odd}\}$ and $D_2 = \{x\colon [-|x|] \text{ is even}\}$. Let $f_1(x) = x + \operatorname{sgn} x$, for $x \in D_1$, and let $f_2(x) = -x + \operatorname{sgn} x$, with $f_2(0) = 0$, for $x \in D_2$. Then $f = f_1 \cup f_2$ satisfies $f(f(x)) = -x$.

II. *F. S. Cater, Portland State University, gives the following generalization.*

THEOREM. *Let a and b be real numbers, $b < 0$. Let F be a function from the real line R into R satisfying Menger's equation*

$$F(f(x)) = a + bx \ (x \in R).$$

Then F has infinitely many discontinuities.

Proof. F is a 1-1 mapping of R onto R because $F \circ F$ is. Then F cannot be everywhere

continuous; for, if it were, F would be monotone and $F \circ F$ would be increasing, impossible. Now we suppose henceforth that F has only a finite number of points of discontinuity. Then the set of points of continuity is an open set with finitely many component intervals $I_1, \dots, I_n$. F is discontinuous at the endpoints of these intervals. On each I_j, F is continuous and monotone, $F(I_j)$ is also an open interval, and F^{-1} is continuous on $F(I_j)$.

We claim that if F is continuous at x, then F is continuous at $F(x)$. Let $y_m \to F(x)$, and say $x \in I_j$, $F(x) \in F(I_j)$. Now $F(I_j)$ is open; so $y_m \in F(I_j)$ for all but finitely many m. But F^{-1} is continuous at $F(x)$, so $F^{-1}(y_m) \to x$ and

$$F(y_m) = F\big(F\big(F^{-1}(y_m)\big)\big) = a + bF^{-1}(y_m) \to a + bx = F\big(F(x)\big).$$

This proves our claim.

Say $F(I_j) \subset I_k$. But every I_k is a subset of the range of F, and there are only finitely many such intervals. Hence $F(I_j) = I_k$, and F defines a permutation q of the intervals $I_1, \dots, I_n$. Likewise F defines a permutation s of the points of discontinuity $u_1, \dots, u_{n-1}$.

We claim that $n > 2$. Suppose that $n = 2$. Say I_1 is unbounded to the left and I_2 is unbounded to the right. Then $F \circ F$ maps I_1 onto I_1 and I_2 onto I_2. Since $b < 0$, this situation is impossible. So there must be an interval I_j of finite length. We may now assume, without loss of generality, that $b = -1$. For if $b \neq -1$, then no two of the intervals

$$I_j, (F \circ F)(I_j), (F \circ F \circ F \circ F)(I_j), \dots$$

would have the same length. Henceforth $F(F(x)) = a - x$ for $x \in R$.

We claim that F is discontinuous at $\frac{1}{2}a$. Assume, to the contrary, that $\frac{1}{2}a \in I_j$. Then $F(I_j) \neq I_j$; for otherwise F would be monotone on I_j and $F \circ F$ would be increasing on I_j, impossible. Say $F(I_j) = I_k \neq I_j$. Then $F(I_k) = I_j$ because $F(F(\frac{1}{2}a)) = \frac{1}{2}a$. And $F(F(I_k)) = F(I_j) = I_k$. But $F(F(I_k)) > \frac{1}{2}a$ $(< \frac{1}{2}a)$ if $I_k < \frac{1}{2}a$ $(> \frac{1}{2}a)$. Thus F is discontinuous at $\frac{1}{2}a$.

Again $F(F(I_j)) > \frac{1}{2}a$ $(< \frac{1}{2}a)$ if $I_j < \frac{1}{2}a$ $(> \frac{1}{2}a)$. Thus q^2 is the product of disjoint 2-cycles and has no fixed point. It follows that q is the product of disjoint 4-cycles and q has no fixed point. Hence 4 divides n.

$F(\frac{1}{2}a) = \frac{1}{2}a$; for otherwise $F(\frac{1}{2}a) = c \neq \frac{1}{2}a$, $F(c) = F(F(\frac{1}{2}a)) = \frac{1}{2}a$, and $F(F(c)) = F(\frac{1}{2}a) = c = a - c$, impossible. So $\frac{1}{2}a$ is the only point fixed by F. Also s^2 is the product of disjoint 2-cycles and fixes exactly one point. It follows that s is the product of disjoint 4-cycles and fixes exactly one point. Since there are $n - 1$ u_j's, it follows that 4 divides $n - 2$. But 4 also divides n. This contradiction completes our proof.

Remarks. There is an F satisfying $F(F(x)) = -x$ $(x \in R)$ whose set of points of discontinuity is discrete. If we delete from R the points ± 1, there is a function F satisfying $F(F(x)) = -x$ $(x \in R)$ that is continuous at all $x \neq 0$. F can follow the pattern

$$(-\infty, -1) \to (0,1) \to (1, \infty) \to (-1, 0) \to (-\infty, -1), \qquad 0 \to 0.$$

Also solved by Miroslav D. Ašić (Yugoslavia), Rufus Isaacs, J. R. Kuttler, L. E. Mattics, E. Triesch (student, Germany), and Richard B. Tucker.

Infinite Series in p-adic Fields

6292 [1980, 225]. *Proposed by L. Van Hamme, Free University of Brussels, Belgium.*

Prove that in any p-adic field, for $p \neq 2$,

$$\sum_{n=1}^{\infty} \frac{n^2 \cdot (n+1)!}{4^{n+1}} = -1.$$

Solution by L. E. Mattics, University of South Alabama. Let $p > 2$ and suppose that K is a

completion of the rationals with respect to $|\ |_p$. Then the series $f(x) = \sum_{n=0}^{\infty} n!x^n$ and all of its derived series converge uniformly for x in K with $|x|_p \leq 1$, and in this disc

$$\frac{d}{dx} x \frac{d}{dx} x f(x) = f'(x) \quad \text{or} \quad x^2 f''(x) = (1-3x)f'(x) - f(x).$$

On the other hand,

$$\begin{aligned} g(x) = \sum_{n=1}^{\infty} n^2(n+1)!x^{n+1} &= x^2 \frac{d}{dx} x \frac{d}{dx} x^{-1}(f(x)-1) \\ &= x^2 f''(x) - xf'(x) + f(x) - 1 \\ &= (1-4x)f'(x) - 1; \end{aligned}$$

hence $g(1/4) = -1$.

Also solved by B. W. Conolly (England), Roger Cuculiere (France), A. A. Jagers (Netherlands), O. P. Lossers (Netherlands), Jose Luis de Miguel (Spain), N. Miku (Netherlands), Brian Peterson, and the proposer.

The proposer adds the following comments.

1. There are very few formulas which are known to be valid in every p-adic field. As far as I know the only formula mentioned in the literature is $\sum_{n=1}^{\infty} n \cdot n! = -1$ (cfr. A. Van Rooij and W. Schikhof, Non-Archimedean Analysis, Nieuw Archief voor Wiskunde (2), 19 (1971) 120–160).
2. One can prove in the same way relations such as

$$\sum_{n=1}^{\infty} n^2 \cdot (n+1)! = 2, \quad \sum_{n=1}^{\infty} n^5 \cdot (n+1)! = 26.$$

3. If we write $S_k = P_k(x)S_0 + Q_k(x)$ then the coefficients of the polynomial $P_k(x)$ are the Stirling numbers of the second kind.

MISCELLANEA

65. It is one thing to learn the science of Analytical Mathematics, and another, to learn its practical applications; and although most of our authors, in this country [Great Britain], have mingled these very different subjects together, it is certainly much to be doubted, whether any advantage has thence accrued either to the student or to the science itself. There is surely a material distinction betwixt the art of Numeration and the processes of Book-keeping. For the understanding of the latter, the former must be learned; but it is quite unnecessary for the mere student of Arithmetic to make himself master of Merchants' Accounts. In the same manner is the general science of Mathematical Analysis related to Geometry and Mechanics. Both these departments of knowledge are greatly facilitated and enlightened by the modern analysis; but the benefit thus conferred is by no means reciprocal, for the principles and methods of the latter are, of themselves, independent, and may be demonstrated without any foreign aid. In Great Britain, however, we do not seem to have sufficiently weighed the importance of this circumstance. Our analytical treatises consist, in great part, of dissertations relative to Statics, Dynamics, & c.; and before the learner can proceed beyond the threshold of the science, his attention is called off to consider the path of a projectile, or the vibrations of a pendulum. It may fairly be asked, what have these subjects to do with analysis? and the only answer that can be given to this question is, that they form some of its numerous applications, and exemplify several of its theories.... On the Continent, Analysis is studied as an independent science. Its general principles are first inculcated; and then the pupil is led to the applications; and the effects have been, that while we have remained nearly stationary during the greater part of the last century, the most valuable improvements have been added to the science in almost every other part of Europe—William Spence, Preface to *An Essay on the Theory of the Various Orders of Logarithmic Transcendents ...*, London and Edinburgh, 1809. (Suggested by R. Askey.)

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, T*(13), S, L. Patterns in Mathematics. Ellery B. Golos. Prindle, 1981, viii + 456 pp. [ISBN: 0-87150-301-8] An introduction to mathematics for college students in general education or the liberal arts. Topics include elementary work in geometry, computation, number theory, probability, statistics and logic. It features a series of introductory preludes containing historical vignettes and many discovery-type exercises. CEC

General, S(13). Computers in Mathematics: A Sourcebook of Ideas. Ed: David H. Ahl. Creative Comp Pr, 1979, x + 213 pp, $15.95 (P). [ISBN: 0-916688-16-X] Collection of articles, activities, and puzzles gleaned largely from Creative Computing Magazine. Covers such diverse topics as computer literacy, problem solving and thinking strategies, computer art, computer assisted instruction, and programming style. Ideas, exercises, projects and problems for senior high students and up. Many topics are valuable without computer access. Good idea generator for future secondary teachers. MW

General, T(13: 1). A Mathematical Panorama: Topics for the Liberal Arts. William P. Berlinghoff, Bodh R. Gulati, Kerry E. Grant. Heath, 1980, xvii + 422 pp, $16.95. [ISBN: 0-669-02423-6] An ambitious course dealing with significant ideas; tilted toward the pure side of mathematics, as opposed to the applied. Ten independent chapters: History, Graph Theory, Matrices, Numbers, Sets, Probability, Statistics, Number Theory, Groups, Axiomatic Systems. LCL

General, L. Basic Library List for Two-Year Colleges, Second Edition. MAA, 1980, 66 pp, $8 (P). [ISBN: 0-88385-436-8] A revision of the 1971 original edition with both additions and deletions designed to keep the list small, to fit even minimal library budgets. Books are arranged in about 20 major sections by topic, with numerous options for local selection. An essential reference for departments and libraries. LAS

General, S*, P, L*. Mathematics Tomorrow. Ed: Lynn Arthur Steen. Springer-Verlag, 1981, vi + 250 pp, $18. [ISBN: 0-387-90564-2] A collection of essays by authors from nearly every nook and cranny where mathematicians are found. These essays are grouped to cover a progression from "What is Mathematics?", through "Teaching and Learning Mathematics," and "Issues of Equality," to "Mathematics Tomorrow." This book provides a basis for a discussion of the near crisis in mathematical and scientific education which is documented in a recent report to the President by the National Science Foundation and the Department of Education. JAS

General, T(13-16). Contemporary Mathematics, Third Edition. Bruce E. Meserve, Max A. Sobel. P-H, 1981, xiii + 670 pp, $19.95. [ISBN: 0-13-170076-6] Written for both liberal arts students and prospective teachers. Covers sets, logic, number systems, number theory, rationals, reals, some geometry, probability and statistics. (First Edition, TR, October 1972; ER, January 1975; Second Edition, TR, June-July 1977.) SG

General, T?(13-14: 1, 2). Mathematics and Statistics for the Bio-Sciences. G. Eason, C.W. Coles, G. Gettinby. Ellis Horwood, 1980, 578 pp, $86.95. [ISBN: 0-85312-175-3] An introduction to calculus, matrix algebra, probability, and statistics with some exercises and examples drawn from the biological sciences. Lack of mathematical rigor will cause some confusion. FLW

General, P, L. An Annotated Bibliography of Films and Videotapes for College Mathematics. David I. Schneider. MAA, 1980, vi + 107 pp, $9 (P). [ISBN: 0-88385-434-1] Several hundred mathematics films, listed by distributor; a catalogue of mathematics TV courses; reference sources; film rental libraries with their mathematics film holdings; subject and title indexes. An invaluable reference for classroom enrichment material. LAS

General, S(10-13), L. Challenging Mathematical Teasers. J.A.H. Hunter. Dover, 1980, 101 pp, $2.75 (P). [ISBN: 0-486-23852-0] 100 new problems whose solutions require clever use of high school algebra, with complete solutions; forty alphametrics with answers; two helpful appendices on Diophantine equations and remainders. A refreshing supplement to challenge bright high school students. LAS

Basic, S(13). Barron's How to Prepare for the Minimum Competency Examination in Mathematics. Angelo Wieland. Barron's Educ Ser, 1981, viii + 344 pp, $6.95 (P). [ISBN: 0-8120-2246-7] Eighty-two lessons, each organized by clearly stated objectives, pertinent definitions and results, model problems, and drill exercises. Four model exams with answers, suggestions and references to appropriate lessons. LCL

Precalculus, T(13: 1). Algebra and Trigonometry. Arnold R. Steffensen, L. Murphy Johnson. Scott F, 1981, 726 pp, $16.95 (P). [ISBN: 0-673-15371-1] Nice worktext. Attractive three-color format with definitions and key theorems highlighted. Numerous examples with explanations of difficult steps in color. Provides hints to difficult exercises and warnings of common misunderstandings and mistakes. Answers immediately follow exercises. Supplementary exercises and tests, an instructor's guide and a computer capable test item bank are available. MW

Precalculus, T(13: 1). College Algebra. Arnold R. Steffensen, L. Murphy Johnson. Scott F, 1981, 512 pp, $15.95 (P). [ISBN: 0-673-15370-3] Contains eleven of the fifteen chapters of College Algebra and Trigonometry, covering all topics except trigonometry. MW

Precalculus, T(13: 1). Intermediate Algebra. Howard A. Silver. P-H, 1981, xv + 478 pp, $18.95. [ISBN: 0-13-469411-2] Skills oriented treatment for students with weak mathematical backgrounds. Emphasis on remembering solution steps, with flow-chart style directions opposite examples and listing of "important procedures" at end of each chapter. MW

Precalculus, T(13: 1). Intermediate Algebra, Third Edition. Margaret L. Lial, Charles D. Miller. Scott F, 1981, xv + 490 pp, $15.95. [ISBN: 0-673-15406-8] Changes from the Second Edition include more work on factoring, wider range of difficulty in exercise sets, and the addition of optional calculator exercises. Increased emphasis on applications; more word problems and many chapters end with a one or two page extended application. (First Edition, TR, October 1972.) MW

Precalculus, T(13: 1). Algebra and Trigonometry, Second Edition. Harley Flanders, Justin J. Price. Saunders Coll Pub, 1981, xii + 524 pp [ISBN: 03-057779-9]; Algebra, Second Edition, 1981, x + 382 pp [ISBN: 03-057801-9]; Precalculus Mathematics, Second Edition, 1981, xii + 587 pp [ISBN: 03-057723-3]. Standard precalculus topics; 75% of text rewritten and some applications added (TR, First Edition, October 1975). Algebra contains chapters 1-6, 10 and 11 of Algebra and Trigonometry (First Edition, TR, November 1975). Precalculus Mathematics is the same as Algebra and Trigonometry except that vectors and conic sections are each a chapter instead of a few sections. LLK

Education, T(13-14: 1). A Problem Solving Approach to Mathematics for Elementary School Teachers. Rick Billstein, Shlomo Libeskind, Johnny W. Lott. Benjamin/Cummings, 1981, xvi + 656 pp, $16.95. [ISBN: 0-8053-0851-2] Fresh approach. Polya's influence is evident throughout. Each chapter begins with a puzzle and concludes with a discussion of its solution. Excellent bibliography and list of typical questions from the classroom for each chapter. Problem solving hints and enough good problems to start a teacher's file. Instructor's guide available. MW

Education, S(13-16). Calculator Explorations and Problems. Don Miller. Cuisenaire Co of America, 1979, 109 pp, $5.50 (P). [ISBN: 0-914040-75-8] Reproducible activities encourage use of calculator as a problem solving aid and data generator. Supplements existing curriculum. Emphasis on searching for patterns, analyzing algorithms and developing estimation skills. Useful for pre-service and in-service courses. MW

Education, S(14-15). Discovery in Mathematics: A Text for Teachers. Robert B. Davis. Cuisenaire Co of America, 1980, 273 pp, $9.50 (P). [ISBN: 0-914040-86-3] New version of Madison Project materials, combining this earlier Student Discussion Guide with the Teacher's Guide. Some new information. Updated introduction and list of references. MW

Education, P. Beiträge zum Mathematikunterricht 1980. Hermann Schroedel, 1980, 372 pp, DM 29,80 (P). [ISBN: 3-507-35016-5] Brief summaries of 95 lectures given at the Fourteenth Annual West German Conference on the teaching of mathematics held in Dortmund in March, 1980. JD-B

History, P, L. Epistemological and Social Problems of the Sciences in the Early Nineteenth Century. Ed: H.N. Jahnke, M. Otte. Reidel Pub, 1981, xlii + 430 pp, $31.50. [ISBN: 90-277-1223-9] Proceedings of a November 1979 workshop held at the Center for Interdisciplinary Research at Bielefeld University. Nearly half of the 22 papers deal with mathematics (e.g., antecedents to Grassmann, origins of pure mathematics, mathematical rigor, probability of judgments). LAS

History, S, P, L. Einstein Between Centuries. Salomon Bochner. Rice University Studies (Vol. 65, No. 3), 1979, iii + 54 pp, (P). [ISBN: 0-89263-242-9] An analysis of Einstein's mode of original thinking, his impact on physics, philosophy and society, and the roots of his resistance to certain more recent innovations--all preceded by a wide-ranging comparative discussion of the intellectual climate of the nineteenth and twentieth centuries before and after Einstein's seminal work. The theme of the two essays in this slim volume is primarily that of continuity vs. discontinuity (in physics, in mathematics and in society at large), with Einstein standing as an interface between them. LAS

Foundations, P. Combinatory Reduction Systems. J.W. Klop. Math. Centre Tracts, No. 127. Math Centrum, 1980, xiii + 317 pp, Dfl. 39,30 (P). [ISBN: 90-6196-200-5] Technical monograph on the syntactical properties of the lambda-calculus, combinatory logic and general term rewriting systems at the borderline of mathematical logic and theoretical computer science. Introduces and develops the general concept of combinatory reduction system together with a generalization of a proof theoretic technique of Nederpelt. GHM

Foundations, T(16-18). Computability and Logic, Second Edition. George Boolos, Richard Jeffrey. Cambridge U Pr, 1980, xi + 285 pp, $39.95; $13.95 (P). [ISBN: 0-521-23479-4; 0-521-29967-5] The

second edition of this well received text is revised and corrected throughout. Intended for a second course in logic it gives excellent coverage of the fundamental theoretical results about logic involving computability, undecidability, axiomatization, definability, incompleteness, etc. Two new chapters: one on modal logic and provability, another on Ramsey's combinatorial theorem. Regrettably there is no mention of the Ramsey-type independence results of Paris and others. Few exercises. (First Edition, TR, March 1975.) GHM

Foundations, T(14-15), L. Set Theory with Applications, Second Edition, Revised and Expanded. Shwu-Yeng T. Lin, You-Feng Lin. Mariner Pub, 1981, ix + 221 pp, $12.95 (P). [ISBN: 0-936166-03-7] This quite elementary treatment of set theory is nonaxiomatic, but lucid and rigorous on an intuitive basis. Covers relations, functions, cardinal and ordinal numbers, and Axiom of Choice, but no construction of number systems. Chapter One presents rudiments of propositional logic, three pages on quantification, and mathematical induction (with Peano's axioms in an appendix). The applications amount to a modest chapter on Boolean algebra and logic circuits. (First Edition, TR, October 1974) GHM

Graph Theory, P, L*. Reviews in Graph Theory. Ed: William G. Brown. AMS, 1980, $200 set (P) [ISBN:0-8218-0214-3]; Volume 1. xvi + 588 pp; Volume 2, xvi + 546 pp; Volume 3, xvi + 574 pp; Volume 4, x + 325 pp. Volumes 1-3 contain approximately 10,000 reviews reprinted from Volumes 1-56 (1940-78) of Math Reviews, arranged in a tree-like classification with citations to as many as seven secondary subject areas. Volume 4 contains author and key indices, and a detailed subject index to the classification scheme. LAS

Combinatorics, S(17), P. Surveys in Combinatorics. Ed: B. Bollobás. London Math. Soc. Lect. Note Ser., No. 38. Cambridge U Pr, 1979, vii + 261 pp, $23.95 (P). [ISBN: 0-521-228468] Proceedings of the Seventh British Combinatorial Conference which took place in Cambridge during August of 1979. The principal speakers review the diverse areas of combinatorics in which they are expert. CEC

Combinatorics, S(16), P*, L. Rudiments of Ramsey Theory. Ronald L. Graham. CBMS Reg. Conf. in Math., No. 45. AMS, 1981, v + 65 pp, $5.60 (P). [ISBN: 0-8218-1696-9] Based on lectures presented at a CBMS regional research conference held at St. Olaf College, June 18-22, 1979. Develops the background for understanding recent developments in Ramsey Theory. Complete proofs are given for most of the basic results. CEC

Number Theory, S*(16-17), P*, L*.** 13 Lectures on Fermat's Last Theorem. Paulo Ribenboim. Springer-Verlag, 1979, xvi + 302 pp, $25.20. [ISBN: 0-387-90432-8] The early history and main theories connected with the problem and analogues of Fermat's Last Theorem. Very readable for mathematicians in general. Gives a good feel for the mathematics without giving the painstaking details. Includes excellent lists of references. CEC

Algebra, S(18), P. Rings with Chain Conditions. A.W. Chatters, C.R. Hajarnavis. Research Notes in Math., No. 44. Pitman Pub, 1980, 197 pp, $18.95 (P). [ISBN: 0-273-08446-1] Account of some of the recent developments in the theory of non-commutative rings with chain conditions. Emphasizes applications of Artinian radical and Goldie's rank function. JRG

Algebra, S(18), P. Set Theoretic Methods in Homological Algebra and Abelian Groups. Paul C. Eklof. Pr U Montreal, 1980, 117 pp, (P). [ISBN: 2-7606-0467-5] Assuming a knowledge of rudimentary set theory and the fundamentals of homological algebra, the theme is to derive the algebraic implications of assuming some of the set-theoretic hypotheses consistent with, but not provable in, ordinary set theory. This includes work of Shelah on Whitehead's problem, Mekler on almost free groups, and the author on modules over Dedekind domains. References. JS

Algebra, P. Integral Representations and Structure of Finite Group Rings. Klaus W. Roggenkamp. Pr U Montreal, 1980, 165 pp, (P). [ISBN: 2-7606-0485-3] A relatively self-contained set of notes on the structure of the group ring (integral, p-adic, or modular) of a finite group. SG

Algebra, P. Representations of Valued Graphs. Vlastiml Dlab. Pr U Montreal, 1980, 190 pp, (P). [ISBN: 2-7606-0503-5] Valued graphs are essentially directed graphs in which a positive integer is associated with every arc. The representation of valued graphs constitutes yet another illustration of how the language and theory of graphs illuminates algebraic theories, e.g., modules (representations of K-algebras), filtered vector spaces (representations of posets), and the classification of linear transformations. This volume--containing a set of lectures given at L'Universite de Montreal--presents the basic ideas of representation theory with the goal of applying it to the theory of the classification of torsion-free abelian groups. SS

Algebra, P. Abelian p-Groups and Mixed Groups. Laszlo Fuchs. Pr U Montreal, 1980, 139 pp, (P). [ISBN: 2-7606-0468-3] Exploration and survey of new results. A basic technique in this study of p-groups is to view the socles as vector spaces over the prime field of characteristic p furnished with valuations given by the height function. A dominating feature of the second half study is to view a mixed group as an extension of a torsion-free group by a torsion group. LCL

Algebra, P. Homological Invariants of Modules Over Commutative Rings. Paul Roberts. Pr U Montreal, 1980, 110 pp, (P). [ISBN: 2-7606-0499-3] Notes based on lectures given at the University of Montreal, summer of 1979. Covers invariants associated with minimal free and minimal injective resolutions, relation between homological dimension and Krull dimension, construction of a minimal free resolution of the ideal of minors of a matrix of indeterminates. KS

Calculus, T(13-14), S, L. Lehrbuch der Analysis, Teil 1. Harro Heuser. Teubner Stuttgart, 1980, 643 pp, DM 48 (P). [ISBN: 3-519-02221-4] First of two volumes. Higher level than standard U.S. calculus texts. Quite readable. Few routine exercises. PH

Calculus, T(13: 2), S. Technical Calculus with Analytic Geometry, Second Edition. Allyn J. Washington. Benjamin/Cummings, 1980, xiii + 514 pp, $18.95. [ISBN: 0-8053-9519-9] Topics covered include basic analytic geometry, differentiation and integration of algebraic and elementary transcendental functions, an introduction to partial derivatives and double integrals, expansion of functions in series, differential equations and a brief introduction to curve fitting. This edition has 40% more exercises and 15% more worked examples than the previous edition. Applications are of the traditional variety. Cookbookish. CEC

Calculus, T(14: 1), S, L*.** Vector Calculus, Second Edition. Jerrold E. Marsden, Anthony J. Tromba. Freeman, 1981, xviii + 591 pp, $22.95. [ISBN: 0-7167-1244-X] In this second edition (First Edition, TR, March 1976; Extended Review, May 1977), optional sections on theoretical topics like the implicit function theorem, properties of the integral and the geometric meaning of divergence and curl have been added. The number of applications and exercises has been increased and numerous historical notes have been added. Now this well-written text is even better. CEC

Calculus, T(13-14: 1-3), S. Calculus, Second Edition. Stanley I. Grossman. Academic Pr, 1981, xviii + 1158 pp, $28.95. [ISBN: 0-12-304360-3] Major changes include earlier introduction to the sine and cosine functions, the distribution of ϵ-δ material over several chapters, a full chapter on analytic geometry, and expansion of the material on vector analysis. In addition, there are hundreds of refinements in response to suggestions from readers. An attractive contender which deserves special consideration. (First Edition, TR, June-July 1979.) LCL

Calculus, T(13-14: 1). Calculus for the Management, Life, and Social Sciences. Bernard Kolman. Acad Pr, 1981, xiii + 514 pp, $19.95. [ISBN: 0-12-417890-1] An intuitive calculus beginning with a lengthy (98 page) review of algebra and analytic geometry and ending with partial differentiation plus a final short chapter on trigonometric functions. Treatment is low-key and fairly standard; one unusual feature is extensive treatment of methods of anti-differentiation before mention of the definite integral. JS

Calculus, T(13-14: 2). Mathematical Analysis for Business and Economics. Jagdish C. Arya, Robin W. Lardner. Prentice-Hall, 1981, 11 + 691 pp, $21.95. [ISBN: 0-13-561019-2]; Applied Calculus for Business and Economics, 1981, xii + 394 pp, $18.95. [ISBN: 0-13-039255-3] In three parts: I--precalculus; II-- finite mathematics, including linear programming and excluding symbolic logic; III--differential and integral calculus of algebraic functions and ln x, e^x--including partial derivatives. Clearly written, with stress on intuitive appeal rather than rigor. Applied Calculus consists almost exactly of Mathematical Analysis, minus the chapters on elementary algebra and finite mathematics. The larger book costs only three dollars more. MB

Real Analysis, S(17-18), P, L. Construction of Borel Measures on Metric Spaces. Peer Kornum. Lect. Notes Ser., No. 54. Aarhus U, 1980, 223 pp, (P). Hausdorff measure perspective. Introduces a construction of "dual" Hausdorff measures called packing measures. Fairly self-contained. Interesting reading. PH

Real Analysis, S(17-18), P. Real Variable Methods in Fourier Analysis. Miguel de Guzmán. Math. Stud., No. 46. North-Holland, 1981, xiii + 392 pp, $44 (P). [ISBN: 0-444-86124-6] Of an introductory character. Well written and properly titled. Unsolved problems discussed in text and listed at the end. PH

Real Analysis, S(17-18), P. A Method of Averaging in the Theory of Orthogonal Series and Some Problems in the Theory of Bases. S.V. Bockarev. Proc. of Steklov Inst. of Math., No. 146. AMS, 1980, vi + 92 pp, $26 (P). [ISBN: 0-8218-3045-7] PH

Real Analysis, T(15-16), S, L. Foundations of Mathematical Analysis. Richard Johnsonbaugh, W.E. Pfaffenberger. Pure and Appl. Math., V. 62. Dekker, 1981, viii + 428 pp, $24.50. [ISBN: 0-8247-6919-8] Fairly standard one-year introduction to real functions. Organization is such that a one-semester course would be confined to a study of limits. First half a bit ponderous. Last three chapters on Inner Product Spaces and Fourier Series, Normed Linear Spaces and the Riesz Representation, the Lebesgue Integral especially well done. Over 750 problems. PH

Real Analysis, T(16-17: 2), S. Mass-und Integrationstheorie: Eine Einführung. Klaus Floret. Teubner Stuttgart, 1981, 360 pp, DM 29,80 (P). [ISBN: 3-519-02059-9] A development of measure and Lebesgue integration theory, along lines suggested by Stone and Daniell. Exercises. JD-B

Real Analysis, P. Proceedings of the Conference on Integration, Topology, and Geometry in Linear Spaces. Ed: William H. Graves. Contemp. Math., V. 2. AMS, 1980, ix + 269 pp, $14 (P). [ISBN: 0-8218-5002-4] 14 papers from a May 1971 conference at the University of North Carolina, dedicated to Billy James Pettis. LAS

Differential Equations, S(18), P. Singular Perturbations and Asymptotics. Ed: Richard E. Meyer, Seymour V. Parter. Academic Pr, 1980, ix + 409 pp, $22. [ISBN: 0-12-493260-6] Collection of invited lectures in honor of Wolfgang Wasow, May 1980, Madison. PH

Differential Equations, S(18), P. Topics in Finite Elasticity. Morton E. Gurtin. CBMS Reg. Conf. in Appl. Math., No. 35. SIAM, 1981, v + 58 pp, $9.50 (P). [ISBN: 0-89871-168-1] Fourteen brief, terse chapters on finite elasticity (theory of elastic materials capable of undergoing large deformations) lead from a treatment of kinematics and stress through the derivation of the basic equations of the theory to a study of the deformations of a cube and a treatment of anti-plane shear. LCL

Differential Equations, P. Admissible Solutions of Hyperbolic Conservation Laws. Tai-Ping Liu. Memoirs No. 240. AMS, 1981, iv + 78 pp, $4.80 (P). [ISBN: 0-8218-2240-3] A study of the regularity, large-time behavior, and the approximation of the solution of the initial value problem. These equations arise in the study of solid mechanics and fluid dynamics with chemicals and multi-phase flows. The analysis is based on the random choice method. LCL

Differential Equations, P. Analysis and Optimisation of Stochastic Systems. Ed: O.L.R. Jacobs, et al. Acad Pr, 1980, xiv + 573 pp, $48. [ISBN: 0-12-378680-0] Refereed papers from the International Conference on Analysis and Optimization of Stochastic Systems held at the University of Oxford, September 6-8, 1978. JAS

Differential Equations, S(16-18), P, L. Sturmian Theory for Ordinary Differential Equations. William T. Reid. Appl. Math. Sci., V. 31. Springer-Verlag, 1980, xv + 559 pp, $28.50 (P). [ISBN: 0-387-90542-1] An historically organized survey of the qualitative analysis of differential equations begun by Sturm in 1836, emphasizing linkages with the calculus of variations. Reid traces the evolution of the theory from Sturm and Liouville through Hilbert to Morse, covering linear equations, boundary value problems and extensions to integral equations, Hilbert space and topology. The manuscript is a typescript written in 1975, just recently prepared for publication by the late author's former students. LAS

Differential Equations, T(18). Smooth Dynamical Systems. M.C. Irwin. Pure and Appl. Math. Academic Pr, 1980, x + 259 pp, $48.50. [ISBN: 0-12-374450-4] Intended to fill the gap between Hirsch and Smale's Differential Equations Dynamical Systems, and Linear Algebra and current research papers in the area. Presumes grounding in several variable differential calculus, some elementary topology, and a little algebraic topology. Watch out for the nine appendixes, ranging from group theory and spectral theory to Liapunov stability and the theory of manifolds. AWR

Numerical Analysis, S(16-18), P, L. Interval Mathematics 1980. Ed: Karl L.E. Nickel. Acad Pr, 1980, xv + 554 pp, $29.50. [ISBN: 0-12-518850-1] Proceedings of an international symposium on this rapidly developing subject which acts as a two-way bridge between mathematics and computing. Topics range from purely theoretical (e.g., set-valued mappings in partially ordered spaces; the importance of 3-valued notions) to important new computational methods (for nonlinear optimization, nonlinear systems of equations, power series expansions, roots of polynomials, eigenvectors, exponentials of matrices) and even computer architecture. LCL

Numerical Analysis, T(17-18: 1, 2). Approximation Theory and Methods. M.J.D. Powell. Cambridge U Pr, 1981, x + 339 pp, $57.50; $19.95 (P). [ISBN: 0-521-22472-1; 0-521-29514-9] A thorough introduction with lots of exercises and a good bibliography. The presentation assumes a background in analysis at the level of Lebesgue measure. The author is most careful to prove all theorems and, even though no programs are presented, he is aware of the interest in computer applications of this material. JAS

Numerical Analysis, T(15-17), S, P, L. Numerical Methods for Stiff Equations and Singular Perturbation Problems. Willard L. Miranker. Math. and Its Appl., V. 5. Reidel, 1981, xiii + 202 pp, $29.95. [ISBN: 90-277-1107-0] A stiff differential equation is an ill-conditioned (nearly singular) initial value problem; a singular perturbation problem is a system of differential equations whose solution is especially sensitive to the value of a small parameter. Both are numerically unstable when solved by traditional computer methods. This concise, clearly-written monograph collects and organizes results from recent literature to address these difficult but very important special cases. LAS

Numerical Analysis, P. Navier-Stokes Equations: Theory and Numerical Analysis, Revised Edition. Roger Temam. Stud. in Math. and Its Appl., V. 2. North-Holland, 1979, x + 519 pp, $68.25; $24.50 (P). [ISBN: 0-444-85307-3; 0-444-85308-1] Only minor changes; a few additional references. (First Edition, TR, December 1978.) LCL

Numerical Analysis, P. The Mathematics of Finite Elements and Applications III: MAFELAP 1978. Ed: J.R. Whiteman. Acad Pr, 1979, xv + 513 pp, $44.50. [ISBN: 0-12-747253-3] Proceedings of the conference held at Brunel University, England, April 18-21, 1978. (TR, I, March 1974; II, November 1980.) JAS

Numerical Analysis, P*. Spline Functions: Basic Theory. Larry L. Schumaker. Wiley, 1981, xiv + 553 pp, $42.50. [ISBN: 0-471-76475-2] A detailed and comprehensive introduction to the theory of polynomial, generalized, and multidimensional splines emphasizing their algebraic, analytic, and approximation-theoretic properties. AO

Numerical Analysis, P. Stable Recursions with Applications to the Numerical Solution of Stiff Systems. J.R. Cash. Acad Pr, 1979, xii + 223 pp, $52. [ISBN: 0-12-163050-1] This monograph focuses on and exploits the relationship between the theories of linear recurrence relations and ordinary

differential equations to produce new iterative schemes for the numerical solution of systems of ordinary differential equations. A number of sections include open problems and are designed to stimulate further research. AO

Numerical Analysis, T(16-17: 1), S, L. Numerical Solution of Partial Differential Equations. Theodor Meis, Ulrich Marcowitz. Appl. Math. Sci., V. 32. Springer-Verlag, 1981, viii + 541 pp, $24 (P). [ISBN: 0-387-90550-2] Three main sections of this text cover initial value problems, boundary value problems, and techniques for the solution of systems of equations. Listings of Fortran programs implementing many of the algorithms are included as is some introductory material on partial differential equations and functional analysis. AO

Numerical Analysis, T(15-16). Basic Numerical Mathematics, V. 1: Numerical Analysis. John Todd. Int. Ser. Num. Math., V. 14. Birkhäuser, 1979, 253 pp, $24. [ISBN: 0-12-692401-5] A tersely written introduction to numerical analysis which presents algorithms for root-finding, interpolation, numerical quadrature, and the solution of ordinary differential equations. Approximately one-third of the book is devoted to detailed solutions of selected exercises. AO

Functional Analysis, P. Funktionalanalysis. A. Göpfert, T. Riedrich. Teubner, 1980, 136 pp, (P). A serious introduction to elementary functional analysis for engineers, economists and other scientists. A number of applications included. Number 22 of a series. PH

Functional Analysis, P. Lecture Notes in Mathematics-826: Les Equations de von Kármán. Philippe G. Ciarlet, Patrick Rabier. Springer-Verlag, 1980, vi + 181 pp, $12.70 (P). [ISBN: 0-387-10248-5] RJA

Functional Analysis, T(17-18: 2), L. Green's Functions and Boundary Value Problems. Ivar Stakgold. Wiley, 1979, xv + 638 pp, $33.95. [ISBN: 0-471-81967-0] A completely rewritten one-volume version of the author's 1967 and 1968 two-volume text Boundary Value Problems of Mathematical Physics (TR, June-July 1967 and November 1968, respectively), featuring more varied and more recent physical applications. Includes introductions to the theory of distributions, metric and Hilbert spaces, operator theory, and integral equations, and their applications to linear and nonlinear problems of mathematical physics. LAS

Optimization, T(14-17: 1, 2), S, L. Quantitative Methods for Decision Making in Business. Richard E. Trueman. Dryden Pr, 1981, 737 pp. [ISBN: 0-03-051356-1] Probability theory, decision making, linear programming, transportation and assignment models, network models, dynamic programming, inventory models, queueing models, Markov analysis, and simulation models. Does not presuppose calculus. FLW

Analysis, S(18), P. Measures of Noncompactness in Banach Spaces. Józef Banaś, Kazimierz Goebel. Lect. Notes in Pure and Appl. Math., V. 60. Dekker, 1980, v + 97 pp, $17.50 (P). [ISBN: 0-8247-1248-X] Designed to introduce this theory to specialists in operator equations. A variety of measures of compactness are discussed. A few applications. PH

Analysis, T*(16-18: 1, 2), S, P, L. Theory and Applications of Fourier Analysis. Charles Sparks Rees, S.M. Shah, C.V. Stanojević. Pure and Appl. Math., No. 59. Dekker, 1981, viii + 419 pp, $37.50. [ISBN: 0-8247-6903-1] Self-contained textbook featuring a careful, well-organized style, complete proofs, many examples and exercises. An introductory chapter on Lebesgue integration makes it accessible to all students with a good understanding of advanced calculus. Topics include trigonometric series, summability, multiple Fourier series, Fourier transform, distributions, orthogonal systems, Bessel functions. LCL

Analysis, T(14-15). Introduction to Applicable Mathematics: Part I, Elementary Analysis. Fred A. Hinchey. Halsted Pr, 1980, vii + 288 pp, $19.95. [ISBN: 0-470-27041-1] A very classical introduction to vector analysis, differential equations, classical special functions, Fourier series, and contour integration. JAS

Analysis, P. Spectral Theory of Ordinary Differential Operators. Erich Müller-Pfeiffer. Trans: M.S.P. Eastham. Ellis Horwood, 1981, 247 pp, $56.95. [ISBN: 0-85312-189-3] Can be read as a sequel to Glazman's Direct Methods of Qualitative Spectral Analysis of Singular Differential Operators. Includes the author's recent results concerning the essential spectrum, the discrete spectrum, and the non-existence of eigenvalues. Note price! LCL

Algebraic Geometry, S*(14-17), L. A Scrapbook of Complex Curve Theory. C. Herbert Clemens. Plenum Pr, 1980, ix + 186 pp, $22.50. [ISBN: 0-306-40536-9]. From the flyleaf: "Eclectic though substantial musings on aspects of the theory of complex algebraic curves." These musings include much inviting material that would be accessible to undergraduates. The book uses only high school algebra at the beginning, but moves gradually to using multivariable analysis and complex variable theory. No exercises, good index, and a few high-powered references. JAS

Algebraic Geometry, P. Moduli, Deformations and Classifications of Compact Complex Manifolds. D. Sundararaman. Research Notes in Math., No. 45. Pitman, 1980, 261 pp, $22.95 (P). [ISBN: 0-273-08458-5] A survey of and introduction to major achievements in the last twenty years. Detailed proofs are replaced by detailed references to original or simplified proofs. As a result the exhaustive bibliography contains a thousand items! JAS

Algebraic Geometry, P. Contributions to Algebraic Geometry: In Honor of Oscar Zariski. Ed: Michael Artin, David Mumford. Johns Hopkins U Pr, 1979, 514 pp, $25. [ISBN: 0-8018-2307-2] A collection of papers in honor of Zariski on his eightieth birthday. JAS

Topology, S(16-18), P. Variable Dimension Fixed Point Algorithms and Triangulations. A.J.J. Talman, G. van der Laan. Math. Centre Tracts, No. 128. Math Centrum, 1980, iv + 172 pp, Dfl. 21 (P). Discusses extensions and applications of the classical fixed point theorems; algorithms to find completely labelled simplices; efficiency of triangulations; variable dimension fixed point algorithms; and geometric interpretations of the algorithms. References. RJA

Topology, P. Odd Primary Infinite Families in Stable Homotopy Theory. Ralph L. Cohen. Memoirs No. 242. AMS, 1981, viii + 92 pp, $5.60 (P). [ISBN: 0-8218-2242-X] Certain elements in the group $Ext_A^{*,*}(Z_p,Z_p)$, where A is the mod p Steenrod algebra for p an odd prime, are shown to survive to the limit in the Adams spectral sequence and to represent stable homotopy elements. JAS

Probability, P. Selected Topics in the Study of Markov Operators. Shaul R. Foguel (Dept. of Math., U. of No. Carolina at Chapel Hill), 1980, 116 pp, $6.50 (P). Concerned with the asymptotic behavior of iterates of general Markov operators, operators whose field automorphism part is trivial, and finally, with Harris operators. TAV

Probability, T(16-17: 2), L. Probability and Measure. Patrick Billingsley. Wiley, 1979, xiv + 515 pp, $26.95. [ISBN: 0-471-03173-9] Probability motivates general theory of measure which motivates random sums, Poisson process, queues, central limit theorem. Nice, well thought out treatment. Written with style. Sufficient problems. PH

Probability, T(16-18: 1, 2), S, P. The Theory of Probability. B.V. Gnedenko. Trans: George Yankovsky. MIR Pub, 1978, 392 pp, $7. Fourth printing; for TR's of previous printings, see February 1968 and August-September 1976. RSK

Probability, P. Lecture Notes in Mathematics-833: Semi-Martingales et Grossissement d'une Filtration. Thierry Jeulin. Springer-Verlag, 1980, ix + 142 pp, $12.70 (P). [ISBN: 0-387-10265-5] A "full account of the subject as it is known at the present time, including a lot of...applications to Markov processes and explicit computations on Brownian motion." JAS

Probability, T, S*(16-18), P, L*. Markov Random Fields and their Applications. Ross Kindermann, J. Laurie Snell. Contemp. Math., V. 1. AMS, 1980, ix + 142 pp, $8.80 (P). [ISBN: 0-8218-5001-6] First volume in new AMS soft-cover, camera-ready series of lecture notes and conference proceedings. This volume provides an introductory survey (for both mathematicians and scientists) of Markov random fields, generalizations of Markov chains that arose from the Ising model of ferromagnetic materials. Basic theory is followed by applications to economics and voting behavior, in which interaction among friends and neighbors (which violates the independence postulates of conventional models) is represented by the Markov random field. LAS

Statistics, T(1, 2). Statistical Inference for Management and Economics, Second Edition. David V. Huntsberger, D. James Croft, Patrick Billingsley. Allyn, 1980, xiv + 690 pp, $17.95. [ISBN: 0-205-06803-0] Introductory course for business and economics students. Treats standard topics in descriptive and inferential statistics, including hypergeometric and Poisson distributions. Also multiple regression, time series and index numbers, and decision theory. Eighteen case studies (average two pages each), with actual business data, emphasize the dynamic environment of realistic statistical problems. GHM

Statistics, T(13: 1, 2). Business and Economic Statistics, Revised Edition. Donald R. Plane, Edward B. Oppermann. Business Pub, 1981, xv + 668 pp, $19.95. [ISBN: 0-256-02438-3] Previous edition published under the title of Statistics for Management Decisions (TR, November 1977). This edition includes about 200 additional exercises, a reorganization of many of the topics, and the inclusion of the Komogorov-Smirnov test for goodness of fit. LCL

Statistics, T(13: 1). Basic Statistics for the Behavioral Sciences. Kenneth Pfeiffer, James N. Olson. HR&W, 1981, xv + 444 pp, $18.95. [ISBN: 0-03-049866-X] Written for the "required" courses. Assuming that learning statistics is like taking medicine, the book uses devices to present concepts as efficiently and painlessly as possible. Each chapter begins with a list of objectives and an overview, and ends with a list of key concepts and procedures. Formulas are given without derivation, but their meaning and necessary assumptions are accurately presented. Includes non-parametric one and two sample tests. Answers to exercises are good model for student exposition. Useful as reference. RJK

Statistics, T(14-15: 1,2). Introductory Statistics for Business and Economics. Iver E. Bradley, John B. South. Dryden Pr, 1981, 610 pp. [ISBN: 0-03-053026-1] Designed for flexibility in both content and level. Presupposes only basic algebra. Technical developments and derivations are found in appendices to most chapters. Primary goal is to give interesting, lucid and simple presentations. Optional chapters on special topics include nonparametrics, decision theory, sampling and quality control. Computer access helpful but not required. GHM

Statistics, T(13: 1). Understanding Basic Statistics. Harvey W. Kushner, Gerald De Maio. Holden-Day, 1980, xiii + 381 pp, $17.95. [ISBN: 0-8162-4874-5] Introduction for students of the social sciences, with many applications drawn from the professional literature. Standard topics (e.g.,

Mendenhall), plus a chapter on multivariate analysis. LCL

Statistics, T(17-18: 1), S, P, L. The Statistical Analysis of Failure Time Data. J.D. Kalbfleisch, R.L. Prentice. Wiley, 1980, xi + 321 pp, $29. [ISBN: 0-471-05519-0] Focuses mainly on regression problems with survival data. "Special attention is paid to problems in the biomedical sciences." Includes computer programs for methods to be used with the proportional hazards model. FLW

Statistics, T(17-18: 1), S, P, L. Robust Statistics. Peter J. Huber. Wiley, 1981, ix + 308 pp, $28.95. [ISBN: 0-471-41805-6] A "systematic exposition of robust statistics." The treatment is theoretical, but the stress is on concepts, rather than mathematical completeness. No exercises. FLW

Computer Programming, S(13), P. Introduction to TI BASIC. Don Inman, Ramon Zamora, Bob Albrecht. Hayden Book, 1980, 305 pp, $10.95 (P). [ISBN: 0-8104-5185-9] An extensive and detailed introduction to TI Basic. You probably need access to a TI 99/4 to appreciate this book. Includes the use of the sound and graphics capabilities of this machine. Lots of examples and some exercises. Easy to read. CEC

Computer Programming, T(13: 1), S. Ten Easy Pieces: Creative Programming For Fun and Profit. Hans Sagan, Carl Meyer, Jr. Hayden Book, 1980, 180 pp, $7.95 (P). [ISBN: 0-8104-5160-3] Basic programming for adults with limited mathematical background through development of computer games and simulations. Readers study the sample run, then copy the program into their own system. Explanations of the reasoning required and the techniques used at each stage. Suggestions for modifications of each program. Readers are alerted to possible differences in dialect. MW

Computer Programming, T(16-18: 1), S, L. Simulation: Principles and Methods. Wayne J. Graybeal, Udo W. Pooch. Winthrop Pub, 1980, xix + 249 pp, $19.95. [ISBN: 0-87626-811-4] Introduction to the general methodology of simulation, including basic aspects of randomness, estimation and inference, experimental design, programming language considerations, verification and validation techniques. Examples and exercises require knowledge of a high-level language (e.g., Fortran). Clear presentation; pleasant format. LCL

Computer Programming, S(15-18), P, L. Software Maintenance Guidebook. Robert L. Glass, Ronald A. Noiseux. P-H, 1981, xi + 193 pp, $21.95. [ISBN: 0-13-821728-9] The phases of the software life cycle are requirements, definition, design, coding, checkout and verification, and maintenance (fixing errors, making user-specified modifications, honing the program to be more useful). This engaging and pioneering work discusses methods and techniques of maintenance, as well as ways of lowering maintenance costs (over half the software development dollar) by better planning, organization, and documentation on the up-front life-cycle phases. LCL

Computer Programming, T(13-18: 1), S. Pascal Programming. Laurence V. Atkinson. Wiley, 1980, x + 428 pp, $49.50. [ISBN:0-471-27773-8] Part one contains programming concepts fundamental to all high level sequential programming languages. The second part focuses on two distinctive features of Pascal, symbolic and subrange types, having no counterpart in other languages. Thirdly, an extensive presentation of the data structuring facilities of Pascal is made. Ideal as an introductory text or as a reference. Many interesting, complete programs. Chapter exercises. Appendexes. Index. RJA

Computer Science, T(17: 1), P. Software Development: A Rigorous Approach. Cliff B. Jones. Prentice-Hall, 1980, xv + 382 pp, $28. [ISBN: 0-13-821884-6] Describes a method for developing software such that each stage of development is supported by a correctness argument. Emphasis on formal notations for algorithms and data types. RJK

Computer Science, P*, L.** A Dictionary of Microcomputing. Philip E. Burton. Garland Pub, 1976, xx + 171 pp, $19.50. [ISBN: 0-8240-9930-3] This knife cuts jargon like cheese. It's extremely strong in hardware and system-level vocabulary and relatively weak in business-oriented software terms. Moderately informal, but well done and thorough. JAS

Computer Science, S(17-18), P. ILP: Intermediate Language for Pictures. P.J.W. ten Hagen, et al. Math. Centre Tracts, No. 130. Math Centrum, 1980, 110 pp, Dfl. 14,70 (P). [ISBN: 90-6196-204-8] This "final" report on a high level interactive graphics language (ILP) "replaces and invalidates" the definition given in a preliminary version by T. Hegen, P.J.W. ten Hagen, et al. in 1977. The language is implemented in Algol 68 and is now working, but not fully documented and packaged for general use. JAS

Computer Science, S(16-18), P, L. Computer Arithmetic in Theory and Practice. Ulrich W. Kulisch, Willard L. Miranker. Comp. Sci. and Appl. Math. Academic Pr, 1981, xiii + 249 pp, $25. [ISBN: 0-12-428650-X]

Computer Science, S(15-17), P*, L.** The Art of Computer Programming, Second Edition. Donald E. Knuth. Seminumerical Algorithms, V. 2. A-W, 1981, xiii + 688 pp, $22.50. [ISBN: 0-201-03822-6] An extensively revised and updated edition of a standard reference work. (First Edition, TR, October 1969; ER, October 1970.) Text discusses the generation and use of sequences of random numbers and algorithms for doing arithmetic operations. AO

Computer Science, T(15-16: 1, 2), S, L*. Elements of the Theory of Computation. Harry R. Lewis, Christos H. Papadimitriou. P-H, 1981, xiv + 466 pp, $22.95. [ISBN: 0-13-273417-6] Introductory text

covering theory of automata and formal languages, computability, computational complexity, mathematical logic. No specific mathematical prerequisites, but maturity required to understand proofs and notation. Good motivation of results and comments on their relation to actual computers and programming. Numerous problems and bibliography at end of each chapter. KS

Computer Science, S(15-18), P, L. Programming Language Standardisation. Ed: I.D. Hill, B.L. Meek. Ellis Horwood, 1980, 261 pp, $71.95. [ISBN: 0-85312-188-5] Aims to dispel confusion and misunderstanding in the programming language standardization process. Part one describes current programming language standards; includes chapters on the process itself, Fortran, Cobol, Algol 60, PL/I, Basic, Pascal, real-time language, data base management systems, graphics, operating system command languages. Part two, on the other hand, discusses issues that might or should affect the standardization process in the future. Appendexes. Index. RJA

Systems Theory, T(17). Introduction to Physical System Modelling. P.E. Wellstead. Acad Pr, 1979, ix + 279 pp, $43. [ISBN: 0-12-744380-0] From the jacket: "The book is designed for M. Sc. courses in automatic control, cybernetics, and engineering system dynamics, and for final year undergraduate courses in systems theory,..." Five case studies are included, and laboratory models are available for teaching aids. AWR

Systems Theory, P. Geometrical Methods for the Theory of Linear Systems. Ed: Christopher I. Byrnes, Clyde F. Martin. Reidel, 1980, ix + 317 pp, $39.50. [ISBN: 90-277-1154-2] Proceedings of the NATO advanced study institute and AMS summer seminar in applied mathematics held at Harvard University, June 18-29, 1979. JAS

Systems Theory, T(17). Stochastic Systems for Management. Winfried K. Grassmann. Elsevier North Holland, 1981, xii + 358 pp, $27.95. [ISBN: 0-444-00449-1] Written with the MBA student in mind, an effort is made to minimize mathematical sophistication. Nevertheless calculus and some probability theory is a prerequisite. Concepts of randomness and stochastic systems are stressed, some background in statistical distributions is presented, much space is devoted to queueing theory, and constant attention is given to such applications as inventory control and facility location. AWR

Applications, S(15-18), P, L. A Guide to Computer Applications in the Humanities. Susan Hockey. Johns Hopkins U Pr, 1980, 248 pp, $16.95. [ISBN: 0-8018-2346-3] Applications include indexes, concordances, dictionaries, vocabulary studies, collocations, dialectology; morphological and syntactic analysis, machine translation; stylistic analysis, authorship studies; textual criticism; sound patterns; indexing, cataloguing, information retrieval. Bibliography. Glossary. Acronyms, abbreviations, program names. Addresses. Index. RJA

Applications, P. Modelling of Dynamical Systems, Volume 1. Ed: H. Nicholson. IEE Con. Eng. Series, V. 12. IEE, 1980, xi + 227 pp, $62. [ISBN: 0-906048-38-9] Seven essays surveying different applications of the modelling process (simulation, chemical processes, refrigeration, nuclear reactors, aerospace systems, marine systems, biological systems), intended to reveal "a certain unit" of representation and behavior for all dynamical systems. LAS

Applications (Astronomy), P. Computational Spherical Astronomy. Laurence G. Taff. Wiley, 1981, x + 233 pp, $28.95. [ISBN: 0-471-06257-X] This book provides an introduction to astrometry as well as a collection of numerical techniques for computations. A large glossary is a useful feature of the book. AO

Applications (Biology), P. Lindenmayer Systems: Structure, Languages, and Growth Functions. P.M.B. Vitányi. Math. Centre Tracts, No. 96. Math Centrum, 1980, vii + 209 pp, Dfl. 25 (P). [ISBN: 90-6196-164-5] A unified treatment of the author's research on the use of mathematical models (L systems) to describe the process of biological development. Requires a rudimentary knowledge of formal language theory; otherwise self-contained. LCL

Applications (Biology), S(16-18), P, L. Mathematics of Cell Electrophysiology. Jane Cronin. Lect. Notes in Pure and Appl. Math., V. 63. Dekker, 1981, viii + 125 pp, $19.75 (P). [ISBN: 0-8247-1157-2] An account of some of the problems in ordinary differential equations which arise in cell electrophysiology, together with a summary of some of the analytic and qualitative techniques which can be used to solve them. The mathematical theory is based primarily on the famous Hodgkin-Huxley work on nerve conduction. Self-contained account of the necessary physiology. LCL

Applications (Biology), T(15-16: 1), S. Introduction to Population Modeling, Second Printing. James C. Frauenthal. Birkhäuser, 1980, xviii + 186 pp, $7.50 (P). [ISBN: 3-7643-3015-5] Republication by Birkhäuser of a UMAP monograph originally published in 1979 (TR, June-July 1979). LAS

Applications (Cybernetics), T(15-16: 1), S, L. Engineering Intelligent Systems: Concepts, Theory, and Applications. Robert M. Glorioso, Fernando C. Colón Osorio. Digital Pr, 1980, xvi + 472 pp, $22. [ISBN: 0-932376-06-1] This is a revision of Engineering Cybernetics which was originally published in 1975. It might serve as an introductory text for a course covering such topics as artificial intelligence, control theory, and robotics aimed at senior engineering students. It is also interesting to scan for replies to the question, "What's it good for?" JAS

Applications (Economics), T(13: 1). Quantitative Business Analysis. William M. Bassin. Bobbs-Merrill Edu Pub, 1981, x + 246 pp, $15.95. [ISBN: 0-672-97696-X] A manual of problem solving techniques for students of business administration. Examples, discussion, and exercises emphasize the

use of mathematical techniques (e.g., basic algebra, linear programming, probability) in analyzing business problems (e.g., inventory management, scheduling problems, break-even analysis, annuities and investments). Assumes only basic algebra; however, optional starred sections and problems requiring matrix algebra or calculus are scattered throughout. LCL

Applications (Engineering), P. The Mathematics of Hydrology and Water Resources. Ed: E.H. Lloyd, T. O'Donnell, J.C. Wilkinson. Academic Pr, 1979, xi + 138 pp, $26. [ISBN: 0-12-453350-7] The principal papers from the conference held at the University of Lancaster in July 1976 with the intention of bringing to the attention of the mathematics community some interesting new problems. JAS

Applications (Engineering), T, P. Applied Symbolic Logic. Edward P. Lynch. Wiley, 1980, xi + 260 pp, $41. [ISBN: 0-471-06256-1] Introduction to Boolean algebra and its applications to process design for mechanical and chemical engineers. First half covers set theory, propositional calculus, simplification of Boolean expressions; second half on logic diagrams for processes and synthesis of fault trees. Few exercises. Material on sets and logic sometimes confusing or wrong, e.g., terms "element" and "subset" not always used properly. KS

Applications (Game Theory), P, L. Control and Dynamic Systems: Advances in Theory and Applications, Volume 17. Ed: C.T. Leondes. Acad Pr, 1981, xviii + 424 pp, $32.50. [ISBN: 0-12-012717-2] Ten expository essays on the theory and applications of differential or difference equations. First introduced by Rufus Isaacs in 1965, differential games have numerous applications to problems of search, evasion and aerial combat. This volume offers a contemporary overview of differential games, stressing real time computer control of dynamic systems. LAS

Applications (Operations Research), P. Techniques in Operational Research, Volume 2: Models, Search and Randomization. Brian Conolly. Ellis Horwood, 1981, 338 pp, $77.95. [ISBN: 0-85312-240-7] A mathematically sophisticated book which aims, for the practitioners of operations research, "to widen the reader's modelling and analytical perspective" by filling gaps the author believes are left by existing texts. Much space is devoted to search for lost objects and to the use of Monte Carlo methods in computation. AWR

Applications (Physics), T(13-14: 1), S, L.** Discovering Relativity for Yourself. Sam Lilley. Cambridge U Pr, 1981, xi + 425 pp, $49.50; $19.95 (P). [ISBN: 0-521-23038-1; 0-521-29780-X] Presuming only arithmetic, this book leads the reader through a thorough but non-technical development of the basic ideas of special and general relativity. It is definitely aimed at serious adults (with math anxiety) who are willing to pursue ideas. It might even be suitable for a math/physics appreciation class, or for special high-school situations. The text proceeds in approximately half page capsules with boldface questions (not exercises) for the reader. In the process algebra and some calculus are developed. A unique, thorough approach for the generalist who wants to be educated rather than trained. JAS

Applications (Physics), S(18), P. Geometric Quantization. Nicholas Woodhouse. Clarendon Pr, 1980, xi + 316 pp, $74. [ISBN: 0-19-853528-7] Assuming some knowledge of coordinate-free differential geometry, quantum mechanics, and Hamiltonian mechanics, the author presents a survey of a mathematical tool which provides relations among classical mechanics, quantum theory, and, to an extent, relativity theory. JAS

Applications (Physics), P. Field Theory, Quantization and Statistical Physics: In Memory of Bernard Jouvet. Ed: E. Tirapegui. Reidel, 1981, xxii + 322 pp, $47.50. [ISBN: 90-277-1128-3] JAS

Applications (Psychology), S? Extra-Sensory Perception of Quarks. Stephen M. Phillips. Theosophical Pub, 1980, xiii + 249 pp, $15. [ISBN: 0-8356-0227-3] A theoretical physicist discusses the relation between recent developments in partical theory and investigations published by the Theosophical Society in 1908 and 1915 (Occult Chemistry) in which atomic substructure was viewed by so called "micro-psi" techniques. JAS

Applications (Social Science), S(17-18), P. Applied Game Theory. Ed: S.J. Brams, A. Schotter, G. Schwödiauer. Physica-Verlag, 1979, 447 pp, (P). [ISBN: 3-7908-0208-5] Proceedings of a conference on applied game theory held at the Institute for Advanced Studies, Vienna, June 13-16, 1978. Twenty-eight papers divided into four parts: Power Analysis, Models and Applications in Political Science, Economics, Control and Confrontation. LCL

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

THE MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

SPRING MEETING OF THE MARYLAND-DISTRICT OF COLUMBIA-VIRGINIA SECTION

The spring meeting of the Maryland-District of Columbia-Virginia Section of MAA was held on Saturday, April 11, 1981 at the College of William and Mary, Williamsburg, Virginia. Eighty-two registrants attended the meeting. *Marcia Sward,* Associate Executive Director of the MAA, gave the invited address. Her topic was "Death on the Highways: Can Mathematics Help?"

Section Chairman *John Smith* presided at the business meeting. *Ernest Mabrey* (Department of Energy) was elected Vice-Chairman for Membership. Other section officers for the coming year are: Chairman, *John Schmeelk* (Virginia Commonwealth University); Immediate Past Chairman, *John Smith* (George Mason University); Vice-Chairman for Programs, *Patrick Hayes* (Federal Reserve); Treasurer, *Arthur Charlesworth* (University of Richmond); AHSME Exam Coordinator, *Edward Bender* (J. Sargeant Reynolds Community College); Newsletter Editor, *Howard Penn* (U. S. Naval Academy); Governor, *Ronald Davis* (Northern Virginia Community College); and Secretary, *Robert Hanson* (James Madison University). Reports on the Annual High School Mathematics Contest, the national newsletter, summer workshops, and future meetings were given. Hearty thanks were expressed to *John Smith* for his leadership as section chairman for the past several years.

The following talks were presented: "Teaching Technical Mathematics," *Ronald Davis,* Northern Virginia Community College; "Computer Use in the Classroom," *James Newsom,* Tidewater Community College; "Teaching Remedial Mathematics," *Betty Weissbecker,* J. Sargeant Reynolds Community College; "Computer Assisted Instruction," *Edward Huff,* Northern Virginia Community College; "Math Labs in Two Year Colleges: An Informal Discussion," *John Massey,* Tidewater Community College; "Least-Cost Livestock Feeding Using Ethanol Byproducts," *Christopher Reed,* TRW, Inc., and *Douglas Samuelson,* Federal Aviation Administration; "An Algorithm to Route Jets for Transporting Checks Among Federal Reserve Banks," *David McCarthy,* Bureau of Labor Statistics and *Y. C. Park,* Naval Sealift Command; "A Model for Scheduling the Production of Currency," *Hosain Ali Mahan,* The George Washington University; "The Applied Mathematics Laboratory at Towson State University," *John Morrison* and *Martha Siegel,* Towson State University; "Recommendations of the CUPM Subpanel on Computing," *George Engel,* Christopher Newport College; "Recommendations of the CUPM Subpanel on Modeling," *Ralph Disney,* Virginia Polytechnic Institute and State University; "Teaching Computer Graphics," *Caren Diefenderfer,* Hollins College; "A Solution for Rubik's Cube," *Howard Penn,* U.S. Naval Academy; "Group Norms and the Grading of Chronologies," *William Wardlow,* U.S. Naval Academy; "On Sets with Distinct Subset Sums," *Paul Stockmeyer,* College of William and Mary; "Discrete Optimization Using Incremental Analysis," *John Drew* and *Margaret Schaefer,* College of William and Mary; "MAA Placement Testing," *George Lowerre,* Northern Virginia Community College (Woodbridge); "Current and Future MAA Activities: An Informal Discussion," *Marcia Sward,* MAA; "Findings of the MAA Committee on Improving Remediation Efforts in the Colleges," *Eleanor Green Jones,* Norfolk State University; "Production and Use of Video Tapes in Developmental Mathematics," *Calude Moore,* Danville Community College.

APRIL MEETING OF THE INDIANA SECTION

The spring meeting of the Indiana Section of the MAA was held at Indiana University-Purdue University at Indianapolis on Saturday, April 11, 1981, with 59 members present. The Indiana Small College Math Competition was held in conjunction with the meeting.

The following papers were presented at the meeting: "Girard Triangles," *Rodney T. Hood,* Franklin College; "A tiling of the Plane with Triangles," *Paul T. Mielke,* Wabash College; "Counting Matrices of Given Rank," *Ralph P. Grimaldi,* Rose-Hulman Institute of Technology; "Let Us Teach Concepts," *Herman Rubin,* Purdue University; "Differential Games," *Leonard D. Berkovitz,* Purdue University; "Computer Graphics in Teaching Mathematics," *Gerald J. Porter,* University of Pennsylvania; "The Mathematics Curriculum in the 80's," *William A. Marion,* Valparaiso University.

At the business meeting led by Chairman *Duane Deal,* the chairman of the nominating committee *Paul Mielke* presented the following slate of officers for 1981-82, which was unanimously approved: Chairman, *M. Mundt,* Valparaiso University; Vice-Chairman, *C. Cowen,* Purdue University; Secretary-Treasurer, *R. Patterson,* Indiana University-Purdue University at Indianapolis. *Meyer Jerison* of Purdue University was elected Governor. Memberships in the MAA were awarded to *Michael Call,* Rose-Hulman Institute of Technology, and *David Dwyer,* Purdue University, in recognition of their performance on the Putnam examination.

MAY MEETING OF THE ROCKY MOUNTAIN SECTION

The sixty-fourth annual meeting of the Rocky Mountain Section of MAA was held on May 1-2, 1981 on the campus of Colorado College, Colorado Springs, Colorado, with 120 members in attendance. Professor *Leonard Gillman,* Treasurer of the MAA, gave the annual banquet address, "We Can't Teach Our Way out of a Paper Bag," and the invited address, "Optimal Strategies in Sports."

The program included three panel discussions: "Employment Opportunities for Students," moderated by Professor *John Hodges* of the University of Colorado; speakers were *Robert Frost,* NOAA; Professor *Kent Goodrich,* University of Colorado; Professor *John Sopka,* Ft. Lewis College; and Professor *S.W. Wilson,* University of Colorado. "How Do Your Start a Computer Science Program?" moderated by *W.S. Dorn* of Denver University. The speakers included Professor *James Davis* of Mesa College; Professor *Charlotte Murphy,* Metro State College; and Professor *Ron Prather* of Denver University. "How Do We Encourage the Exceptional Student?" moderated by Professor *David Ballew* of South Dakota School of Mines and Technology. The participants were Professor *Stephen Bronn,* University of Southern Colorado and Professor *Paul Perlmutter* of Colorado College.

Twenty-three papers were contributed, eight of them by undergraduate students. "Some Unorthodox Thoughts About Quantum Mechanics," by *Jan Mycielski,* University of Colorado, Boulder; "Can a Mathematician Find Happiness in the Computing Field?" by *William Marion,* Valparaiso University; "Eigenvalues

of Tridiagonal Matrices," by *Dale M. Rognlie*, South Dakota School of Mines and Technology; "A Critique of Axiomatic Systems," by *George S. Donovan*, Metro State College; "Unsolved Mathematical Problems Arising from Spectrum Management Issues," by *William K. Hale*, US Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunications Sciences, Boulder; "A Recent Result in Continued Fraction Theory," by *John P. Gill*, University of Southern Colorado; "Some Elementary Models in Business and Economics Part I," by *C. G. Mendez*, Metro State College; "It's not What *You Say*, It's What *They Do* (A Brief Report on a Modularized Approach to Calculus)," by *William R. Astle* and *Thomas E. Kelley*, Colorado School of Mines; "Extremals of $\int L dt$ and of $\int F(L)dt$." by *Roger Opp* and *Ron Weger*, South Dakota School of Mines and Technology; "An Elementary Proof That $(1 + 1/n)^n$ Tends to e," by *Lee Badger*, Ft. Lewis College; "The Evolution of Calculus Teaching and Texts," by *Wolfgang Thron*, University of Colorado, Boulder; "Pancakes and Permutations," by *Leslie E. Shader*, University of Wyoming; "The Evolution of a Generalization," by *Richard Gibbs*, Ft. Lewis College; "The Pi Theorem of Buckingham--Dimensional Analysis and Some Examples," by *Robert S. Fisk*, Colorado School of Mines; "How to Cut a Triangle," by *Alexander Soifer*, University of Colorado, Colorado Springs.

Papers presented by undergraduate students: "Analysis of a Trisection of the Angle Algorithm" by *Leon Nelson*, South Dakota School of Mines and Technology; "Robots--Are They Here to Stay?" by *Janet Potts*, South Dakota School of Mines and Technology; "The Fibonacci Sequence and Evolution," by *Terri Bush*, Ft. Lewis College; "An Application of Game Theory in Taking Tests," by *Dean Mogck*, South Dakota School of Mines and Technology; "The Theory of Evolution Applied to Programming Language," by *Colleen Quatier*, South Dakota School of Mines and Technology; "Computer Graphics of Parametric Equations," by *Gary Ricard*, South Dakota School of Mines and Technology; "Monte Carlo Methods for Testing Large Primes," by *Anthony Wakely*, South Dakota School of Mines and Technology; "The Relations of Differentiable Functions and the Power Series," by *Brian Bunsness*, South Dakota School of Mines and Technology.

Professor *William Ramaley* of Ft. Lewis College, Chairperson of the Section, presided at the annual business meeting. The new officers elected for 1981-82 are as follows: Chairperson, Prof. *John Gill*, University of Southern Colorado; Chairperson-elect, Prof. *George Donovan*, Metro State College; Vice-Chairperson, Prof. *Aubrey Owen*, Community College of Denver; Program Chairman, Prof. *Gale Nash*, Western State College; Secretary-Treasurer, Prof. *David Ballew*, SD School of Mines and Technology. The section heard the minutes, treasurer's report, Prof. *Duane Porter's* Governor's report, and a report from Professor *Gillman* on "super-section" meetings. The section recognized Prof. *Dean C. Benson* of SD School of Mines and Technology and Prof. *Ray Hanna* of the University of Wyoming on the occasion of their retirement and commended them for their long service to the Section.

SPRING 1981 MEETING OF THE NORTH CENTRAL SECTION

The spring meeting of the North Central Section was held at Mankato State University, Mankato, Minnesota, on May 1 and 2. Section President, *Stephen Hilding* of Gustavus Adolphus College, presided at the business meeting. Items of business included the presentation of a citation by Governor *Dale Varbert* to *A. Wayne Roberts* of Macalester College and to *John Sitcomb* of the University of North Dakota for their long service as coordinators of the MAA high school exams in Minnesota and North Dakota. the 21 students from the North Central Section scoring in the top 500 on the 1980 Putnam Exam were also honored. New officers elected were: *Sabra Anderson*, University of Minnesota, Duluth, Chairperson-elect; *Alan Kirch*, Macalester College, Secretary-treasurer; and *James Rue*, University of North Dakota, member-at-large.

Invited addresses were given by *Steve Galovich* of Carleton College, "Arithmetic in Characteristic p" and by Chairman-elect of the MAA, *R. D. Anderson*, who spoke on "Algorithmically Defined Functions."

Contributed papers were: "The Arithmetic of Convex Polytopes" by *Walter Sizer* of Moorhead State University; "Probability Distribution Fitting for Large Loss Insurance Data" by *Bruce Binzel, Kari Johnson, Jennifer Smith, Karen Thompson, Carol Veum*, St. Olaf College (students); "Parameter Estimation in Linear Combinations of Exponentials" by *Dale Larson, Kathryn Lenz, Barry Mason, Richard Selby*, St. Olaf College (students); "A Note on Factoring Differential Equations" by *Clayton Knoshaug*, Bemidji State University; "Field Structures on the Natural Numbers" by *Jon Shreve*, Carleton College (student); "An Interdisciplinary Course Based on Godel, Escher, Bach" by *H. B. Coonce*, Mankato State University; "Squares and Non-Squares in Finite Fields" by *Roger Avelsgaard*, Bemidji State University; "Graphing Non-Real Branches of Garden Variety Equations" by *Ron Rietz*, Gustavus Adolphus College.

ALLEGHENY MOUNTAIN SECTION MEETING

The Allegheny Mountain Section of MAA met May 15-16, 1981 at Duquesne University, Pittsburgh, PA. Invited lectures were given by *George E. Andrews* of Penn State University, "More on Ramanujan's Lost Notebook," and by *Alfred B. Willcox*, Executive Director, MAA, "Some Bridges to and From Mathematics or There is a Mathematician Loose in the Supermarket."

There was a panel discussion of "Job Opportunities in Mathematics." Another panel consisting of *Donald Platte*, Mercyhurst College, Coordinator; *James E. Allison*, Bethany College; *Charles Cable*, Allegheny College; and *Richard MacCamy*, Carnegie-Mellon University, discussed "Trends in College Mathematics Curricula."

Four undergraduate students gave talks: *Dave Plottell* of Allegheny College on "Desarguesian and Non-Desarguesian Projective Planes;" *Don Kocher* of Allegheny College on "The Projective Plane and Incidence Matrices" *Theresa Presecan McChesney* of Westminster College on "Optimal Control;" and *Dan Duh* of Allegheny College on "Optimal Harvesting and Dynamic Programming."

Contributed papers by faculty members were "Micro-computers as an Instructional Aid in Mathematics," *Roy Meyers* and *Richard Reynolds*, Penn State University at New Kensington and McKeesport; "Improving the Learning Strategies of the Non-Major," *Mary Kay Hudspeth*, Penn State University at University Park; "Perturbation Solutions of Non-Linear Difference Equations," *Richard F. Melka*, University of Pittsburgh at Bradford; "On the Asymptotic Behavior of Solutions of Ordinary Differential Equations," *M. M. Subramaniam*, Penn State University, Delaware County Campus.

CALENDAR OF FUTURE MEETINGS

Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.
Sixty-second Summer Meeting, Toronto, Canada, August 23–25, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Allegheny College, Meadville, Pennsylvania, April 1982.

EASTERN PENNSYLVANIA AND DELAWARE, Villanova University, Pennsylvania, November 21, 1981.

FLORIDA, Valencia Community College, Orlando, March 5–6, 1982.

ILLINOIS, Southern Illinois University, Edwardsville, April 30–May 1, 1982.

INDIANA

INTERMOUNTAIN

IOWA, Grinnell College, Grinnell, March 26–27, 1982.

KANSAS, Emporia State University, Emporia, April 2–3, 1982.

KENTUCKY, University of Kentucky, Lexington, April 2–3, 1982.

LOUISIANA–MISSISSIPPI, University of Southwestern Louisiana, Lafayette, February 12–13, 1982.

MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, George Washington University, Washington, D.C., November 14–15, 1981.

METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.

MICHIGAN, first Friday and Saturday in May. Deadline for papers six weeks before meeting.

MISSOURI, University of Missouri, Rolla, April 9–10, 1982.

NEBRASKA, Kearney State College, Kearney, April 2–3, 1982.

NEW JERSEY, Trenton State College, Trenton, fall 1981.

NORTH CENTRAL, end of October and April. Deadline for papers October 1 and April 1.

NORTHEASTERN, Trinity College, Hartford, Connecticut, November 20–21, 1981.

NORTHERN CALIFORNIA, University of California, Davis, February 20, 1982.

OHIO, Capital University, Columbus, April 30–May 1, 1982.

OKLAHOMA–ARKANSAS, University of Arkansas, Fayetteville, March 25–27, 1982.

PACIFIC NORTHWEST, Western Washington University, Bellingham, Washington, June 17–19, 1982.

ROCKY MOUNTAIN, Western State College, Gunnison, Colorado, April 30–May 1, 1982.

SEAWAY, SUNY, College at Brockport, Brockport, New York, November 6–7, 1981.

SOUTHEASTERN, Emory University, Atlanta, Georgia, April 1982.

SOUTHERN CALIFORNIA, University of California, Santa Barbara, November 13–14, 1981.

SOUTHWESTERN, University of Arizona, Tucson, April 1982.

TEXAS, Lamar University, Beaumont, April 9–10, 1982.

WISCONSIN, University of Wisconsin, Fond du Lac, late March 1982.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.

AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES, Las Vegas, Nevada, November 11–14, 1982.

AMERICAN MATHEMATICAL SOCIETY, Cincinnati, Ohio, January 13–16, 1982.

AMERICAN SOCIETY FOR ENGINEERING EDUCATION

ASSOCIATION FOR COMPUTING MACHINERY, Los Angeles, California, November 9–11, 1981.

ASSOCIATION FOR SYMBOLIC LOGIC, Philadelphia, Pennsylvania, December 1981.

ASSOCIATION FOR WOMEN IN MATHEMATICS

CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES

FIBONACCI ASSOCIATION

INSTITUTE OF MATHEMATICAL STATISTICS

MU ALPHA THETA

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, Canada, April 14–17, 1982.

OPERATIONS RESEARCH SOCIETY OF AMERICA, Detroit Plaza, Detroit, Michigan, April 19–21, 1982.

PI MU EPSILON

SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION, Carrousel Inn, Columbus, Ohio, November 5–6, 1981.

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Stanford University, Stanford, California, July 19–23, 1982 (30th Anniversary Meeting).

Eminent Mathematicians and Mathematical Expositors speak to

STUDENTS and TEACHERS in...

The NEW MATHEMATICAL LIBRARY

An internationally acclaimed paperback series providing.

- stimulating excursions for students beyond traditional school mathematics
- supplementary reading for school and college classrooms
- valuable background reading for teachers
- challenging problems for solvers of all ages from high school competitions in the US and abroad

The New Mathematical Library is published by the MATHEMATICAL ASSOCIATION OF AMERICA. The volumes are paperbound.

For information regarding the price of these publications, please contact The Mathematical Association of America at the address listed below.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven, NML-01

WHAT IS CALCULUS ABOUT? By W. W. Sawyer, NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman, NML-03

GEOMETRIC INEQUALITIES, By N. D. Kazarinoff, NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Contests sponsored by the MAA, NCTM, Mu Alpha Theta, The Society of Actuaries, and the Casualty Actuarial Society. Covers the period 1950-1960. Compiled and with solutions by C. T. Salkind. NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis, NML-06

USES OF INFINITY, by Leo Zippin, NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields, NML-08

CONTINUED FRACTIONS, by C. D. Olds, NML-09

GRAPHS AND THEIR USES, by Oystein Ore, NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eötvos Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport, NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, by A. Aaboe, NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus, NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven, NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs, NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Contests for the period 1961-1965. NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod, NML-18

GEOMETRY REVISITED, by H.S.M. Coxeter, and S. L. Greitzer, NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore, NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields, NML-21

ELEMENTARY CRYPTANALYSIS — A Mathematical Approach. by Abraham Sinkov, NML-22

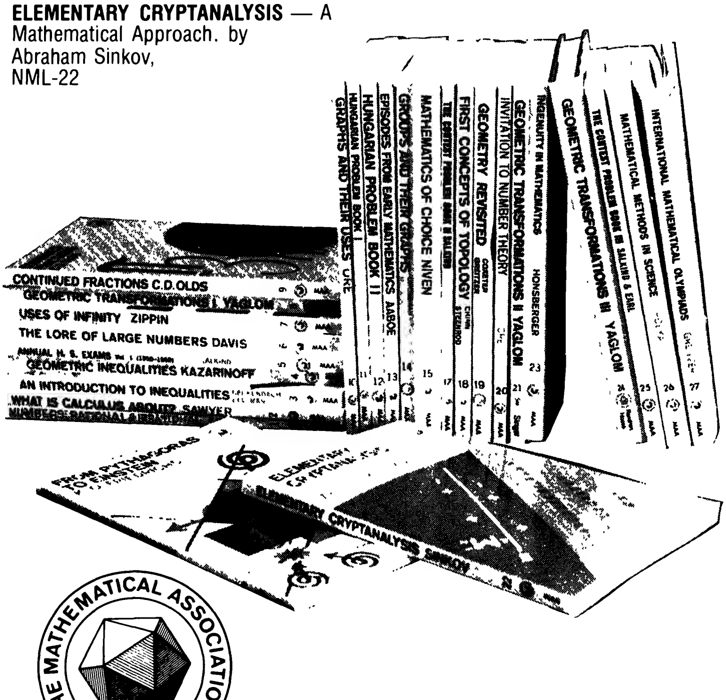
INGENUITY IN MATHEMATICS, by Ross Honsberger, NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer, NML-24

THE CONTEST PROBLEM BOOK III. A continuation of NML-05 and NML-17, containing problems and solutions from the Annual High School Mathematics Contests for the period 1966-1972. NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Pólya, NML-26

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977. Problems, with solutions, from the first nineteen International Mathematical Olympiads. Compiled and with solutions by S. L. Greitzer. NML-27



THE MATHEMATICAL ASSOCIATION OF AMERICA

Send orders to: **The Mathematical Association of America**
1529 Eighteenth St., N.W., Washington, D.C. 20036

The Mathematical Association of America
CORPORATE MEMBERS

The Mathematical Association of America gratefully acknowledges the interest and support of the following:

CORPORATE MEMBERS

Addison-Wesley Publishing Company, Inc.
American Mathematical Society, Inc.
Birkhäuser Boston, Inc.
International Business Machines Corporation
Pergamon Press, Inc.
Saunders College Publishing, Inc.

CONTRIBUTING CORPORATE MEMBER

Bell Laboratories

U.S. POSTAL SERVICE
STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION
(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: THE AMERICAN MATHEMATICAL MONTHLY
A. PUBLICATION NO: A M M Y A E
2. DATE OF FILING: Sept. 14, 1981
3. FREQUENCY OF ISSUE: monthly except July and August
A. NO. OF ISSUES PUBLISHED ANNUALLY: Ten
B. ANNUAL SUBSCRIPTION PRICE: Library $36.00, Indiv. Member $14.00
4. COMPLETE MAILING ADDRESS OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*: 1529 18th Street, NW, Washington, DC 20036
5. COMPLETE MAILING ADDRESS OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: 1529 18th Street, NW, Washington, DC 20036
6. FULL NAMES AND COMPLETE MAILING ADDRESS OF PUBLISHER, EDITOR, AND MANAGING EDITOR *(This item MUST NOT be blank)*
PUBLISHER *(Name and Complete Mailing Address)*: The Mathematical Association of America, 1529 18th St., NW, Washington, DC 20036
EDITOR *(Name and Complete Mailing Address)*: R. P. Boas, Northwestern University, Dept. of Mathematics, Evanston, IL 60201
MANAGING EDITOR *(Name and Complete Mailing Address)*: Dr. Raoul Hailpern, SUNY at Buffalo, Buffalo, NY 14214
7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed)*

FULL NAME	COMPLETE MAILING ADDRESS
The Mathematical Association of America, Inc.	1529 18th St., NW Washington, DC 20036

NAME	COMPLETE MAILING ADDRESS
None	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 411.3, DMM only)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*
(1) [X] HAS NOT CHANGED DURING PRECEDING 12 MONTHS
(2) [ ] HAS CHANGED DURING PRECEDING 12 MONTHS
(If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES *(Net Press Run)*	20,670	20,500
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	none	none
2. MAIL SUBSCRIPTION	18,278	17,919
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	18,278	17,919
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY AND OTHER FREE COPIES	54	54
E. TOTAL DISTRIBUTION *(Sum of C and D)*	18,332	17,973
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	2,338	2,527
2. RETURN FROM NEWS AGENTS	none	none
G. TOTAL *(Sum of E, F1 and 2 — should equal net press run shown in A)*	20,670	20,500

11. I certify that the statements made by me above are correct and complete
SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Alfred B. Willcox

PS Form 3526 (Page 1) June 1980 *(See instruction on reverse)*

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.

INDEX TO VOLUME 88, 1981
THE AMERICAN MATHEMATICAL MONTHLY

AUTHOR INDEX

SUBJECT INDEX

This index uses the current version of AMS 1980 Mathematics Subject Classification. Underlined entries indicate that the subject heading is secondary.

MISCELLANEA

PROBLEMS AND SOLUTIONS

PROBLEMS PROPOSED

PROBLEMS SOLVED

SOLUTIONS

Numbers in boldface type refer to problems; those in lightface, to pages.

INDEX OF SPECIAL RECURRING TOPICS

TELEGRAPHIC REVIEWS

REVIEWS

Names of authors are in ordinary type; those of reviewers, in capitals.

FILMS

NEWS AND NOTICES

PERSONAL ITEMS

Necrology

GENERAL INFORMATION

MATHEMATICAL ASSOCIATION OF AMERICA

REPORTS AND ANNOUNCEMENTS OF THE ASSOCIATION AND ITS SECTIONS

Meetings and Announcements of the Association

Meetings of the Sections

DECEMBER

THE AMERICAN MATHEMATICAL MONTHLY

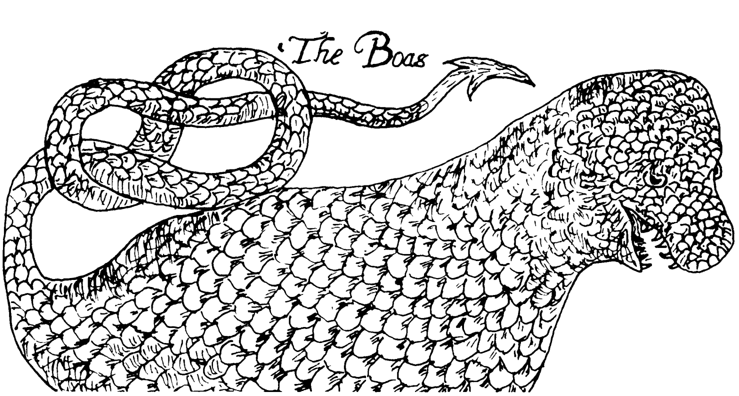
Volume 88, Number 10

Can Mathematics Be Intelligible?

Permanents — An Old Conjecture
(Egorychev's Proof, Presented by Knuth)

Tiling by Unequal Triangles

Hyperbolic Geometry



This picture, including the label, is from Edward Topsell's History of Serpents, London, 1658.

Report on Unsolved Problems

Detailed contents on cover 4

1981

THE AMERICAN MATHEMATICAL MONTHLY

(FOUNDED IN 1894 BY BENJAMIN F. FINKEL)

AN OFFICIAL JOURNAL OF
THE MATHEMATICAL ASSOCIATION OF AMERICA

VOLUME 88 NUMBER 10

(ISSN 0002-9890)

NOTICE TO AUTHORS

Research papers per se are unsuitable; see statement of policy (Vol. 85, p. 1).

Please follow the format in current issues of this MONTHLY. Manuscripts must be legibly typewritten or reproduced from typewritten copy, double spaced with wide margins, and on only one side of the paper. Three copies should be submitted to the appropriate editor and one kept by the author as protection against loss. The author's *full* mailing address must appear at the beginning of the paper, in a line following the author's name. All authors are asked to supply subject classifications with their manuscripts, if possible from the "1980" classification that appears in the 1980 and later Mathematical Reviews Indexes.

Estimated median time from submission to publication: Main Articles 15 months, Progress Reports 6 months, Math. Notes 15 months, Unsolved Problems 6 months, Classroom Notes 12 months, Math. Education 10 months.

MAIN ARTICLES: to P. R. HALMOS, Department of Mathematics, Indiana University, Bloomington, IN 47405; NOTES, etc.: to the corresponding Associate Editor; EDITORIAL CORRESPONDENCE (about articles submitted before May 1981) and REPRINT PERMISSION: to R. P. BOAS, Department of Mathematics, Northwestern University, Evanston, IL 60201 (see also the copyright notice below); ADVERTISING CORRESPONDENCE: to RAOUL HAILPERN, Mathematical Association of America, SUNY at Buffalo, Buffalo, NY 14214; CHANGE OF ADDRESS and SUBSCRIPTIONS: to A. B. WILLCOX, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036; BACK ISSUES: Contact P. and H. Bliss Co., Middletown, CT 06457.

The annual subscription price for the American Mathematical Monthly to an individual member of the Association is $14, included as part of the annual dues of $30. Students receive a 50% discount. The library subscription price is $36 per year.

PUBLISHED BY THE ASSOCIATION at Washington, D.C., and Montpelier, Vermont, during the months of January, February, March, April, May, June-July, August-September, October, November, December.

Second-class postage paid at Washington, D.C., and additional mailing offices.

CAN WE MAKE MATHEMATICS INTELLIGIBLE?

R.P. BOAS
2440 Simpson St., Evanston, IL 60201

Why is it that we mathematicians have such a hard time making ourselves understood? Many people have negative feelings about mathematics, which they blame, rightly or wrongly, on their teachers [**1**]. Students complain that they cannot understand their textbooks; they have been doing this ever since I was a student, and presumably for much longer than that. Professionals in other disciplines feel compelled to write their own accounts of the mathematics they had trouble with. However, it was not until after I became editor of this MONTHLY that I quite realized how hard it is for mathematicians to write so as to be understood even by other mathematicians (outside of fellow specialists). The number of manuscripts rejected, not for mathematical deficiencies but for general lack of intelligibility, has been shocking. One of my predecessors had much the same experience 35 years earlier [**2**].

To put it another way, why do we speak and write about mathematics in ways that interfere so dramatically with what we ostensibly want to accomplish? I wish I knew. However, I can at least point out some principles that are frequently violated by teachers and authors. Perhaps they are violated because they contradict what many of my contemporaries seem to consider to be self-evident truths. (They also have little in common with the MAA report on how to teach mathematics [**3**].)

Abstract Definitions. Suppose you want to teach the "cat" concept to a very young child. Do you explain that a cat is a relatively small, primarily carnivorous mammal with retractile claws, a distinctive sonic output, etc.? I'll bet not. You probably show the kid a lot of different cats, saying "kitty" each time, until it gets the idea. To put it more generally, generalizations are best made by abstraction from experience. They should come one at a time; too many at once overload the circuits.

There is a test for identifying some of the future professional mathematicians at an early age. These are students who instantly comprehend a sentence beginning "Let X be an ordered quintuple $(a, T, \pi, \sigma, \mathfrak{B})$, where . . ." They are even more promising if they add, "I never really understood it before." Not all professional mathematicians are like this, of course; but you can hardly succeed in becoming a professional unless you can at least understand this kind of writing.

However, unless you are extraordinarily lucky, most of your audience will not be professional mathematicians, will have no intention of becoming professional mathematicians, and will never become professional mathematicians. To begin with, they won't understand anything that starts off with an abstract definition (let alone with a dozen at once), because they don't yet have anything to generalize from. Please don't immediately write me angry letters explaining how important abstraction and generalization are for the development of mathematics: I *know* that. I also am sure that when Banach wrote down the axioms for a Banach space he had a lot of specific spaces in mind as models. Besides, I am discussing only the communication of mathematics, not its creation.

For example, if you are going to explain to an average class how to find the distance from a point to a plane, you should first find the distance from $(2, -3, 1)$ to $x - 2y - 4z + 7 = 0$. After that, the general procedure will be almost obvious. Textbooks used to be written that way. It is a good general principle that, if you have made your presentation twice as concrete as you think you should, you have made it at most half as concrete as you ought to.

Remember that *you* have been associating with mathematicians for years and years. By this

The author is Professor Emeritus of Mathematics, Northwestern University. The present issue of this MONTHLY is the last under his editorship.

time you probably not only think like a mathematician but imagine that everybody thinks like a mathematician. Any nonmathematician can tell you differently.

Analogy. Sometimes your audience will understand a new concept better if you explain that it is similar to a more familiar concept. Sometimes this device is a flop. It depends on how well the audience understands the analogous thing. An integral is a limit of a sum; therefore, since sums are simpler (no limiting processes!), students will understand how integrals behave by analogy with how sums behave. Won't they? In practice, they don't seem to. Integrals are simpler than sums for many people, and there may be some deep reason for this [**4**].

Vocabulary. Never introduce terminology unnecessarily [**5**]. If you are going to have to mention a countable intersection of open sets—just once!—there is no justification for defining G_δ's and F_σ's.

I have been assured that nobody can really understand systems of linear equations without all the special terminology of modern linear algebra. If you believe this you must have forgotten that people understood systems of linear equations quite well for many years before the modern terminology had been invented. The terminology allows concise statements; but concision is not the alpha and omega of clear exposition. Modern terminology also lets one say more than could be said in old-fashioned presentations. Nevertheless, at the beginning of the subject a lot of the students' effort has to go into memorizing *words* when it could more advantageously go into learning mathematics. Paying more attention to vocabulary than to content obscures the content. This is what leads some students to think that the real difference between Riemann and Lebesgue integration is that in one case you divide up the x-axis and in the other you divide up the y-axis.

If you think you can invent better words than those that are currently in use, you are undoubtedly right. However, you are rather unlikely to get many people except your own students to accept your terminology; and it is unkind to make it hard for your students to understand anyone else's writing. One Bourbaki per century produces about all the neologisms that the mathematical community can absorb.

In any case, if you *must* create new words, you can at least take the trouble to verify that they are not already in use with different meanings. It has not helped communication that "distribution" now means different things in probability and in functional analysis. On the other hand, if you need to use old but unfashionable words it is a good idea to explain what they mean. A friend of mine was rebuked by a naïve referee for "inventing" bizarre words that had actually been invented by Kepler.

It is especially dangerous to assume either that the audience understands your vocabulary already or that the words mean the same to everybody else that they do to you. I know someone who thinks that everybody from high school on up knows all about Fourier transforms, in spite of considerable evidence to the contrary. Other people think that everybody knows what they mean by Abel's theorem, and therefore never say which of Abel's many theorems they are appealing to.

An even more serious problem comes from what (if it didn't violate my principles) I would call geratologisms: that is, words and phrases that, if not actually obsolete in ordinary discourse, are becoming so. Contemporary prose style is simpler and more direct than the style of the nineteenth century—except in textbooks of mathematics. While I was writing this article I was teaching from a calculus book that begins a problem with, "The strength of a beam varies directly as . . ." I do not know whether the jargon of variation is still used in high schools, but in any case it isn't learned: only one student in a class of 45 had any idea what the book meant (and he was a foreigner). Blame the students if you will, blame the high schools; for my own part I blame the authors of the textbook for not realizing that contemporary students speak a different language. Another current calculus book says, "Particulate matter concentrations in parts per million theoretically decrease by an inverse square law." You couldn't get away with that in *Newsweek* or even in *The New Yorker*, but in a textbook . . .

Authors of textbooks (lecturers, too) need to remember that they are supposed to be addressing the students, not the teachers. What is a function? The textbook wants you to say something like, "a rule which associates to each real number a uniquely specified real number," which certainly defines a function—but hardly in a way that students will comprehend. The point that "a definition is satisfactory only if the students understand it" was already made by Poincaré **[6]** in 1909, but teachers of mathematics seem not to have paid much attention to it.

The difficulties of a vocabulary are not peculiar to mathematics; similar difficulties are what makes it so frustrating to try to talk to physicians or lawyers. They too insist on a rich technical language because "it is so much more precise that way." So it is, but the refined terminology is clearer only when rigorous distinctions are absolutely necessary. There is no use in emphasizing refined distinctions until the audience knows enough to see that they are needed.

Symbolism is a special kind of terminology. Mathematics can't get along without it. A good deal of progress has depended on the invention of appropriate symbolism. But let's not become so fascinated by the symbols that we forget what they stand for. Our audience (whether it is listening or reading) is going to be less familiar with the symbolism than we are. Hence it is not a good idea (to take a simple example) to say "Let f belong to L^2" instead of "Let f be a measurable function whose square is integrable," unless you are sure that the audience already understands the symbolism. Moreover, if you are not actually going to use L^2 as a Hilbert space, but want only the properties of its elements as functions, the structure of the space is irrelevant and calling attention to it is a form of showing off—mild, but it *is* showing off. If the audience doesn't know the symbolism, it is mystified; if it does know, it will be wondering when you are going to get to the point.

My advice about new terminology applies with even greater force to new symbolism. Do not create new symbolism, or change the old, unnecessarily; and admit (if necessary) that usage varies and explain the existing equivalences. If your $\Phi(x)$ also appears in the literature as $P(x)$ or $P(x)+\frac{1}{2}$ or $F(x)$, *say so*. Irresponsible improvements in notation have already caused enough trouble. I don't know who first thought of using θ in spherical coordinates to mean azimuth instead of colatitude, as it almost universally did and still does in physics and in advanced mathematics. It's superficially a reasonable convention because it makes θ the same as in plane polar coordinates; however, since r is different anyway, that isn't much help. The result is that students who go beyond calculus have to learn all the formulas over again. Such complications don't bother the true-blue pure mathematicians, those who would just as soon see Newton's second law of motion stated as $\mathbf{v} = (d/d\sigma)(\mathfrak{R}\mathbf{q})$, but they do bother many students, besides irritating physical scientists.

Proofs. Only professional mathematicians learn anything from proofs. Other people learn from explanations. I'm not sure that even mathematicians learn much from proofs in fields with which they are not familiar. A great deal can be accomplished with arguments that fall short of being formal proofs. I have known a professor (I hesitate to say "teacher") to spend an entire semester on a proof of Cauchy's integral theorem under very general hypotheses. A collection of special cases and examples would have carried more conviction and left time for more varied and interesting material, besides leaving the audience better equipped to understand, apply, generalize, and teach Cauchy's theorem.

I cannot remember who first remarked that a sweater is what a child puts on when its parent feels cold; but a proof is what students have to listen to when the teacher feels shaky about a theorem. It has been claimed [7] that "some of the most important results . . . are so surprising at first sight that nothing short of a proof can make them credible." There are fewer of these than you think.

Experienced parents realize that when a child says "Why?" it doesn't necessarily want to hear a reason; it just wants more conversation. The same principle applies when a class asks for a proof.

Rigor. This is often confused with generality or completeness. In spite of what reviewers are likely to say, there is nothing unrigorous in stating a special case of a theorem instead of the most general case you know, or a simple sufficient condition rather than a complicated one. For example, I prefer to give beginners Dirichlet's test for the convergence of a Fourier series: "piecewise monotonic and bounded" is more comprehensible than "bounded variation"; and, in fact, equally useful after one more theorem (learned later).

The compulsion to tell everything you know is one of the worst enemies of effective communication. We mathematicians would get along better with the Physics Department if, for example, we could bring ourselves to admit that, although their students need some Fourier analysis for quantum mechanics, they don't need a whole semester's worth—two weeks is nearer the mark.

Being more thorough than necessary is closely allied to **pedantry**, which (my dictionary says) is "excessive emphasis of trivial details."

Here's an example. Suppose students are looking for a local minimum of a differentiable function f, and they find critical points at $x = 2$, $x = 5$, and nowhere else. Suppose also that they do not want to use (or are told not to use) the second derivative. Some textbooks will tell them to check $f(2 + h)$ and $f(2 - h)$ for all small h. Students naturally prefer to check $f(3)$ and $f(1)$. The pedantic teacher says, "No"; the honest teacher admits that any point up to the next critical point will do.

Enthusiasm. Teachers are often urged to show enthusiasm for their subjects. Did you ever have to listen to a really enthusiastic specialist holding forth on something that you did not know and did not want to know anything about, say the bronze coinage of Poldavia in the twelfth century or "the doctrine of the enclitic *De*" [**8**]? Well, then.

Skills. A great deal of the mathematics that many mathematicians support themselves by teaching consists of subjects like elementary algebra or calculus or numerical analysis—skills, in short. It is not always easy to tell whether a student has acquired a skill or, as we like to put it, "really" learned a subject. The difficulty is much like that of deciding whether apes can use language in a linguistically interesting way or whether they have just become very clever at pushing buttons and waving their hands [**9**]. Mathematical skills are like any other kind. If you are learning to play the piano, you usually start by practicing under supervision; you don't begin with theoretical lectures on acoustical vibrations and the internal structure of the instrument. Similarly for mathematical skills. We often read or hear arguments about the relative merits of lectures and discussions, as if these were the only two ways to conduct a class. Having students practice under supervision is another and very effective way. Unfortunately it is both untraditional and expensive.

Even research in mathematics is, to a considerable extent, a teachable skill. A student of G. H. Hardy's once described to me how it was done. If you were a student of Hardy's, he gave you a problem that he was sure you could solve. You solved it. Then he asked you to generalize it in a specific way. You did that. Then he suggested another generalization, and so on. After a certain number of iterations, you were finding (and solving) your own problems. You didn't necessarily learn to be a second Gauss that way, but you could learn to do useful work.

Lectures. These are great for arousing the emotions. As a means of instruction, they ought to have become obsolete when the printing press was invented. We had a second chance when the Xerox machine was invented, but we seem to have muffed it. If you *have* to lecture, you can at least hand out copies of what you said (or wish you had said). I know mathematicians who contend that only through their lectures can they communicate their personal attitudes toward their subjects. This may be true at an advanced level, for pre-professional students. Otherwise I wonder whether these mathematicians' personalities are really worth learning about, and (if so) whether the students couldn't learn them better some other way (over coffee in the cafeteria, for example.)

One of the great mysteries is: How can people manage to extract useful information from incomprehensible nonsense? In fact, we can and do. Read, for example, in Morris Kline's book [10] about the history of the teaching of calculus. Perhaps this talent that we have can explain the popularity of lectures. One incomprehensible lecture is not enough, but a whole course may be effective in a way that one incomprehensible book never can. I still contend that a comprehensible book is even better.

Conclusion. I used to advise neophyte teachers: "Think of what your teachers did that you particularly disliked—and don't do it." This was good advice as far as it went, but it didn't go far enough. My tentative answer to the question in my title is, "Yes; but don't be guided by introspection." You cannot expect to communicate effectively (whether in the classroom or in writing) unless and until you understand your audience. This is not an easy lesson to learn.

References

1. See, for example, Sydney J. Harris, column for February 9, 1980, Chicago Sun-Times and elsewhere.
2. L. R. Ford, Retrospect, this MONTHLY, 53 (1946) 582–585.
3. College Mathematics: Suggestions on How to Teach It, Mathematical Association of America, 1972.
4. D. R. Stoutemyer, Symbolic computation comes of age, SIAM News, 12, no. 6 (December 1979) 1, 2, 9.
5. The same point has been made by P. R. Halmos in How to Write Mathematics, L'Enseignement mathématique, (2) 16 (1970) 123–152.
6. H. Poincaré, Science et méthode, 1909, Book II, Chapter 2.
7. H. and B. S. Jeffreys, Methods of Mathematical Physics, 2nd ed., Cambridge University Press, 1950, p. v.
8. Robert Browning, "A Grammarian's Funeral," in The Complete Poetic and Dramatic Works of Robert Browning, Houghton Mifflin, Boston and New York, 1895, pp. 279–280.
9. For example, E. S. Savage-Rumbaugh, D. M. Rumbaugh, and S. Boysen, Do apes use language? Amer. Scientist, 68 (1980) 49–61.
10. Morris Kline, Mathematics: The Loss of Certainty, Oxford University Press, New York, 1980.

A PERMANENT INEQUALITY

DONALD E. KNUTH
Computer Science Department, Stanford University, Stanford, CA 94305

A famous problem that has resisted the attacks of many of the world's greatest mathematicians was resolved in 1980 by G. P. Egorychev [5], who proved a conjecture that B. L. van der Waerden had made in 1926 [15]. The conjecture, which is now a theorem, states that the permanent of an $n \times n$ doubly stochastic matrix is never less than $n!/n^n$; the latter value, which is obtained when all entries of the matrix are equal to $1/n$, is therefore the minimum. The purpose of this note is to give an essentially self-contained exposition of Egorychev's proof and the auxiliary results that preceded it, using only elementary concepts of mathematics (except at one point).*

1. Introduction to Quadratic Forms. Our discussion will be based mostly on facts about matrices and quadratic forms that we will prove "from scratch." A *quadratic form* $f(x_1, \ldots, x_n)$ of n variables is an expression

$$f(x_1, \ldots, x_n) = \sum_{i,j} f_{ij} x_i x_j \tag{1.1}$$

defined by some $n \times n$ matrix of real coefficients f_{ij}. The sum is over $1 \leq i, j \leq n$, which we abbreviate to "i, j". Since the coefficient of $x_i x_j$ in $f(x_1, \ldots, x_n)$ is $f_{ij} + f_{ji}$ when $i \neq j$, we can

The author is Professor of Computer Science at Stanford. He is a member both of the National Academy of Sciences and of the National Academy of Engineering and is a recipient of the National Medal of Science.

The name conventionally transliterated Egorychev is pronounced (approximately) YeGORycheff.—*Editors*

***Added in proof:** See the additional information on p. 798, this issue.

assume without loss of generality that $f_{ij} = f_{ji}$ = one-half of the coefficient of $x_i x_j$; then the matrix of coefficients is symmetric about its diagonal. For example, the quadratic form $x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$ can be specified either by the triangular matrix

$$\begin{pmatrix} 1 & -2 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

or by the symmetric matrix

$$\begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix};$$

we will always assume that it corresponds to the latter.

The most successful way to deal with quadratic forms seems to be to consider what happens when we do linear transformations of the variables, namely to replace x_i by $\Sigma_j t_{ij} y_j$ for $1 \leqslant i \leqslant n$; then $f(x_1, \ldots, x_n)$ is transformed into another quadratic form $g(y_1, \ldots, y_n)$, whose matrix of coefficients has $g_{ij} = \Sigma_{k,l} f_{kl} t_{ki} t_{lj}$. If the transformation matrix (t_{ij}) is nonsingular, so that we can invert it and express the y's in terms of the x's, the quadratic forms f and g are called *equivalent*.

We shall now prove that every quadratic form is equivalent in this sense to a quadratic form of a very simple type. The idea is to use a transformation that can be thought of as a highly generalized version of "completing the square":

LEMMA 1.1. *Let $f(x_1, \ldots, x_n) = \Sigma_{i,j} f_{ij} x_i x_j$ be a quadratic form and let the vector $(a_1, \ldots, a_n)$ of real numbers be such that $a_1 \neq 0$ and $f(a_1, \ldots, a_n) = c \neq 0$. Then the nonsingular transformation defined by*

$$x_i = a_i \Big(y_1 - \sum_{j \geqslant 2} y_j \sum_k f_{jk} a_k / c \Big) + y_i \cdot (i \geqslant 2), \tag{1.2}$$

$$y_1 = \sum_{i,j} f_{ij} a_i x_j / c, \qquad y_i = x_i - x_1 a_i / a_1 \quad \text{for } i \geqslant 2, \tag{1.3}$$

makes $f(x_1, \ldots, x_n) = cy_1^2 + g(y_2, \ldots, y_n)$, where g is a quadratic form in $n - 1$ variables. (The notation '$(i \geqslant 2)$' in (1.2) denotes 1 if $i \geqslant 2$ and 0 otherwise; it is often convenient to use relations as expressions in this way.)

Proof. It is not difficult to verify that (1.2) and (1.3) are inverses of each other. Let t_{ij} denote the coefficient of y_j in x_i. The coefficient of y_1^2 in $f(x_1, \ldots, x_n)$ is $\Sigma_{i,j} f_{ij} t_{i1} t_{j1} = \Sigma f_{ij} a_i a_j = c$; and the coefficient of $y_1 y_k$ for $k \geqslant 2$ is

$$\sum_{i,j} f_{ij}(t_{i1}t_{jk} + t_{j1}t_{ik}) = \sum_{i,j} f_{ij} \left(a_i \Big((j = k) - \frac{a_j}{c} \sum_l f_{kl} a_l \Big) + a_j \Big((i = k) - \frac{a_i}{c} \sum_l f_{kl} a_l \Big) \right),$$

which nicely cancels to zero. ■

LEMMA 1.2. *Every quadratic form $f(x_1, \ldots, x_n)$ is equivalent to a simple quadratic form*

$$g(y_1, \ldots, y_n) = y_1^2 + \cdots + y_p^2 - y_{p+1}^2 - \cdots - y_r^2 \tag{1.4}$$

for some $0 \leqslant p \leqslant r \leqslant n$.

Proof. If $f(a_1, \ldots, a_n)$ is zero for all $(a_1, \ldots, a_n)$, we have form (1.4) with $p = r = 0$. If $f(a_1, \ldots, a_n)$ is nonzero for at least one vector $(a_1, \ldots, a_n)$, then some $a_i \neq 0$; by permuting variables if necessary we can assume that $a_1 \neq 0$. Now we use the construction of Lemma 1.1, but with y_1 replaced by $z = y_1 |c|^{1/2}$, thereby obtaining $f(x_1, \ldots, x_n) = \pm z^2 + g(y_2, \ldots, y_n)$. The result follows by induction. ■

As an example of the process used in these proofs, let us find a reduced form equivalent to $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$. First, $f(1,0,0) = 1$, so the transformation of Lemma 1.1 applies with $x_1 = y_1 + y_2 + y_3$, $x_2 = y_2$, $x_3 = y_3$; we have $f(x_1, x_2, x_3) = y_1^2 - 4y_2y_3$. Let $g(y_2, y_3) = -4y_2y_3$; since $g(1,1) = -4$, we can let $y_2 = \frac{1}{2}z_1 - \frac{1}{2}z_2$, $y_3 = \frac{1}{2}z_1 + \frac{1}{2}z_2$, so that $-4y_2y_3 = -z_1^2 + z_2^2$. Thus we have

$$f(x_1, x_2, x_3) = y_1^2 + z_2^2 - z_1^2 = (x_1 - x_2 - x_3)^2 + (x_3 - x_2)^2 - (x_2 + x_3)^2.$$

The method of reduction in Lemma 1.1 was essentially discovered by Lagrange in 1759; the next basic fact that we need is "Sylvester's law of inertia," which dates from 1852.

LEMMA 1.3. *The numbers p and r of Lemma* 1.2 *are unique*; *in other words*, *if we have equivalent quadratic forms*

$$y_1^2 + \cdots + y_p^2 - y_{p+1}^2 - \cdots - y_r^2 = z_1^2 + \cdots + z_q^2 - z_{q+1}^2 - \cdots - z_s^2, \tag{1.5}$$

then $p = q$ and $r = s$.

Proof. Let $y_i = \Sigma t_{ij}z_j$, and suppose that $p < q$. Then we can find values $z_1^*, \ldots, z_q^*$ not all zero such that $\Sigma_{1 \leq j \leq q} t_{ij}z_j^* = 0$ for $1 \leq i \leq p$, since this is a system of p homogeneous linear equations in $q > p$ unknowns z_j^*. Let $z_{q+1}^* = \cdots = z_n^* = 0$ and $y_i^* = \Sigma_{1 \leq j \leq n} t_{ij}z_j^*$ for all i. Then (1.5) reduces to

$$-y_{p+1}^{*2} - \cdots - y_r^{*2} = z_1^{*2} + \cdots + z_q^{*2},$$

which can only hold if $z_1^* = \cdots = z_q^* = 0$, a contradiction. Hence $p = q$. Now if $r < s$, we can similarly find $z_{q+1}^*, \ldots, z_s^*$ not all zero such that $\Sigma_{q < j \leq s} t_{ij}z_j^* = 0$ for $p < i \leq r$; setting $z_1^* = \cdots = z_q^* = z_{s+1}^* = \cdots = z_n^* = 0$, we get

$$y_1^{*2} + \cdots + y_p^{*2} = -z_{q+1}^{*2} - \cdots - z_s^{*2},$$

another contradiction. ■

As a consequence of Lemma 1.3 we can refer to $p(f)$ and $r(f)$ as well-defined invariants of the quadratic form f.

We will need a somewhat more subtle fact later:

LEMMA 1.4. *Let $f_\theta(x_1, \ldots, x_n)$ be the quadratic form*

$$f_\theta(x_1, \ldots, x_n) = (1 - \theta)f_0(x_1, \ldots, x_n) + \theta f_1(x_1, \ldots, x_n) \tag{1.6}$$

that changes from f_0 to f_1 as θ varies from 0 *to* 1. *If $r(f_\theta) = n$ for $0 \leq \theta \leq 1$, then $p(f_0) = p(f_1)$.*

This lemma can be proved by using well-known facts of analysis since $p(f_\theta)$ is the number of positive eigenvalues of the real symmetric matrix f_θ, and since $r(f_\theta) = n$ is equivalent to saying that f_θ has no zeroes as eigenvalues; the result follows because the roots of a polynomial are continuous functions of their coefficients, and the eigenvalues of f_θ are roots of polynomials whose coefficients are continuous functions of θ. A proof closer to first principles can be formulated by showing that for all θ_0 in [0, 1] there exists $\epsilon > \theta$ such that $p(f_\theta)$ is constant for $|\theta - \theta_0| < \epsilon$; by compactness, we can then cover [0, 1] with a finite number of open intervals of constant p. The existence of ϵ follows from a proof like that of Lemma 1.3; details are left to the interested reader. ■

2. Quadratic Forms and Permanents. The *permanent* of an $n \times n$ matrix $A = (a_{ij})$ is the sum

$$\mathrm{per}(A) = \sum_\pi a_{1\pi(1)} \cdots a_{n\pi(n)}, \tag{2.1}$$

taken over all permutations $\pi = \pi(1) \ldots \pi(n)$ of $\{1, \ldots, n\}$. We will write a_i for the ith row $(a_{i1}, \ldots, a_{in})$ of the matrix A, and $\mathrm{per}(A) = \mathrm{per}(a_1, \ldots, a_n)$ if we are enumerating its rows.

The next part of the theory is a proof of two lemmas, which are proved simultaneously by induction on n (i.e., first we prove both of them for $n = 2$, then both for $n = 3$, and so on).

LEMMA 2.1. *Let $a_1, \ldots, a_{n-1}$ be vectors of nonnegative numbers in which at least $n+1-i$ elements of a_i are positive, and suppose that $b = (b_1, \ldots, b_n)$ is any vector of real numbers such that*

$$\operatorname{per}(a_1, \ldots, a_{n-1}, b) = 0. \tag{2.2}$$

Then

$$\operatorname{per}(a_1, \ldots, a_{n-2}, b, b) \leq 0; \tag{2.3}$$

furthermore, $\operatorname{per}(a_1, \ldots, a_{n-2}, b, b) = 0$ *if and only if* $b_1 = \cdots = b_n = 0$.

LEMMA 2.2. *Let $a_1, \ldots, a_{n-2}$ be as in Lemma* 2.1, *and let f be the quadratic form*

$$f(x_1, \ldots, x_n) = \operatorname{per}(a_1, \ldots, a_{n-2}, x, x) \tag{2.4}$$

where x stands for the vector $(x_1, \ldots, x_n)$. Then $r(f) = n$ and $p(f) = 1$.

Proof. Both results are clear for $n = 2$: If $a_{11} > 0$ and $a_{12} > 0$ and $a_{11}b_2 + a_{12}b_1 = 0$ then $2b_1b_2 \leq 0$; and in this case we have $2b_1b_2 = 0$ if and only if $b_1 = b_2 = 0$. Furthermore $2x_1x_2 = ((x_1 + x_2)/\sqrt{2})^2 - ((x_1 - x_2)/\sqrt{2})^2$. So we shall assume that $n \geq 3$ and that both lemmas have been proved for $n - 1$.

In the quadratic form (2.4), the value of f_{ij} is the permanent of the $(n-2) \times (n-2)$ matrix obtained by removing columns i and j of the matrix $(a_1, \ldots, a_{n-2})$, if $i \neq j$; also $f_{ii} = 0$. If $r(f) < n$, the matrix f is singular, so there is a nonzero vector $(c_1, \ldots, c_n)$ such that $\Sigma_j f_{ij} c_j = 0$ for all i. This is equivalent to saying that $\operatorname{per}(a_1, \ldots, a_{n-2}, c, x) = 0$ for all x, because of our interpretation of f_{ij}; in particular, $\operatorname{per}(a_1, \ldots, a_{n-2}, c, c) = 0$. Furthermore we have $\operatorname{per}_j(a_1, \ldots, a_{n-2}, c) = 0$ for all j, where per_j denotes the permanent obtained by removing column j. By induction, $\operatorname{per}_j(a_1, \ldots, a_{n-3}, c, c) \leq 0$ for all j. Now

$$0 = \operatorname{per}(a_1, \ldots, a_{n-2}, c, c) = \sum_i a_{(n-2)j} \operatorname{per}_j(a_1, \ldots, a_{n-3}, c, c) \leq 0;$$

hence we have $\operatorname{per}_j(a_1, \ldots, c, c) = 0$ whenever $a_{(n-2)j} > 0$. This occurs for at least two values of j, in fact for at least three; hence $c_1 = \cdots = c_n = 0$, a contradiction.

We have proved half of Lemma 2.2, the fact that $r(f) = n$. For the other half, it suffices by Lemma 1.4 to compute $p(f)$ in the special case that $a_1 = \cdots = a_{n-2} = (1, 1, \ldots, 1)$, since we can transform the rows one by one from this case into (2.4), and all the intermediate quadratic forms have rank n by what we have already proved. In this special case the quadratic form is $(n-2)!$ times the special quadratic form defined by $f_{ij} = (i \neq j)$. When $n = 4$, for example, the special form is $2(x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4)$. The general transformation method of Lemma 1.2 shows that this form can be represented for example as

$$\sum_{i \neq j} x_i x_j = (n-1)s^2 - (x_2 - s)^2 - \cdots - (x_n - s)^2, \tag{2.5}$$

where $s = x_1 + \frac{1}{2}(x_2 + \cdots + x_n)$. Hence $p(f) = 1$.

We turn now to the proof of Lemma 2.1. The hypotheses on $a_1, \ldots, a_{n-1}$ imply that $\operatorname{per}(a_1, \ldots, a_{n-1}, a_{n-1}) > 0$; hence we have $f(a_{(n-1)1}, \ldots, a_{(n-1)n}) = c > 0$, in terms of the quadratic form (2.4). We may assume without loss of generality that $a_{(n-1)1} > 0$, by permutation of columns. Therefore if we apply the construction of Lemma 1.1 we obtain

$$f(x_1, \ldots, x_n) = cy_1^2 + g(y_2, \ldots, y_n).$$

By Lemma 2.2, $g(y_2, \ldots, y_n) \leq 0$ for all $(y_2, \ldots, y_n)$, and it is zero if and only if $y_2 = \cdots = y_n$. We have

$$y_1 = \sum_{1 \leq i, j \leq n} f_{ij} a_{(n-1)i} x_j / c = \operatorname{per}(a_1, \ldots, a_{n-1}, x)/c$$

by (1.2); this, together with our hypothesis (2.2), implies that $f(b_1,\ldots,b_n) \leq 0$. Finally, $f(b_1,\ldots,b_n) = 0$ implies that we have $y_1 = \cdots = y_n = 0$ when $(x_1,\ldots,x_n) = (b_1,\ldots,b_n)$; hence $b_1 = \cdots = b_n = 0$. ■

Now we come to the main theorem on which Egorychev's proof of van der Waerden's conjecture rests. This theorem is essentially due to A. D. Aleksandrov, who published it in 1938 [**1**], using a more general framework that was not obviously related to permanents. Aleksandrov's motivation was the study of the volume of n-dimensional convex sets, for which a similar geometric inequality had been deduced by W. Fenchel in 1936 [**7**]. A search through *Science Citation Index* for 1960–1980 reveals that the geometric aspects of Aleksandrov's work became well known, but the algebraic aspects were rarely cited in the Western world. The only exceptions are Busemann's book [**3**], which paraphrases part of Aleksandrov's proof, and Schneider's paper [**12**], which presents an alternative derivation. (See also the recent work by Teissier [**14**] and Stanley [**13**].) Egorychev discovered that Aleksandrov's work was relevant to permanents after studying a variety of new formulas for the permanent function [**6**].

THEOREM 2.3. *Let $a_1,\ldots,a_{n-1}$ be nonnegative vectors such that a_i contains at least $n+1-i$ positive entries, and let a_n be any vector of real numbers. Then*

$$\operatorname{per}(a_1,\ldots,a_{n-1},a_n)^2 \geq \operatorname{per}(a_1,\ldots,a_{n-2},a_{n-1},a_{n-1})\operatorname{per}(a_1,\ldots,a_{n-2},a_n,a_n), \tag{2.6}$$

and equality holds if and only if $a_n = \lambda a_{n-1}$ for some real number λ.

Proof. Let $\operatorname{per}(a_1,\ldots,a_{n-1},a_n) = \lambda \operatorname{per}(a_1,\ldots,a_{n-1},a_{n-1})$; then λ is well defined since $\operatorname{per}(a_1,\ldots,a_{n-1},a_{n-1}) > 0$. If we set $b = a_n - \lambda a_{n-1}$, we have (2.2), since the permanent is a linear function of each of its rows. Hence Lemma 2.1 tells us that

$$\begin{aligned}
0 &\geq \operatorname{per}(a_1,\ldots,a_{n-2},b,b)\\
&= \operatorname{per}(a_1,\ldots,a_{n-2},b,a_n) - \lambda\operatorname{per}(a_1,\ldots,a_{n-2},b,a_{n-1})\\
&= \operatorname{per}(a_1,\ldots,a_{n-2},a_n,a_n) - 2\lambda\operatorname{per}(a_1,\ldots,a_n)\\
&\quad +\lambda^2\operatorname{per}(a_1,\ldots,a_{n-2},a_{n-1},a_{n-1})\\
&= \operatorname{per}(a_1,\ldots,a_{n-2},a_n,a_n) - \lambda^2\operatorname{per}(a_1,\ldots,a_{n-2},a_{n-1},a_{n-1})\\
&= \operatorname{per}(a_1,\ldots,a_{n-2},a_n,a_n) - \frac{\operatorname{per}(a_1,\ldots,a_n)^2}{\operatorname{per}(a_1,\ldots,a_{n-2},a_{n-1},a_{n-1})}.
\end{aligned}$$

Equality holds if and only if $b = 0$, i.e., $a_n = \lambda a_{n-1}$. ■

COROLLARY 2.4. *Let $a_1,\ldots,a_{n-1}$ be nonnegative vectors and let a_n be arbitrary. Then the inequality (2.6) holds.*

Proof. Consider the vectors $a_i + (\epsilon,\ldots,\epsilon)$ and take the limit as ϵ approaches zero through positive values. ■

Incidentally, the example

$$\operatorname{per}^2\begin{pmatrix}1&1&0\\0&1&1\\1&0&1\end{pmatrix} = \operatorname{per}\begin{pmatrix}1&1&0\\0&1&1\\0&1&1\end{pmatrix}\operatorname{per}\begin{pmatrix}1&1&0\\1&0&1\\1&0&1\end{pmatrix}$$

shows that some condition on positivity is necessary to obtain Theorem 2.3's strong statement about the conditions of equality in (2.6).

3. Doubly Stochastic Matrices. The matrix (a_{ij}) is called *doubly stochastic* if its elements are

nonnegative and if all row sums and column sums are equal to 1. Thus, each row can be regarded as a probability distribution, and so can each column; it is for this reason that the matrix is "stochastic" in a double sense.

Doubly stochastic matrices have many pleasant properties that make them important in applications. For example, it is easy to verify that the product AB of two doubly stochastic matrices A and B is doubly stochastic; and if $0 \leqslant \theta \leqslant 1$, the convex combination $(1-\theta)A + \theta B$ is also doubly stochastic. If we imagine a network of n points with a_{ij} units of flow proceeding from point i to point j, the total flow in and out of each point is equal to unity.

The simplest kind of doubly stochastic matrix is a *permutation matrix*, in which all a_{ij} are 0 or 1; there is exactly one 1 in every row and every column. If $\pi(1) \ldots \pi(n)$ is a permutation of $\{1, \ldots, n\}$, we let P_π be the corresponding permutation matrix (p_{ij}), where $p_{ij} = (j = \pi(i))$. Garrett Birkhoff proved in 1946 [**2**] that every doubly stochastic matrix is a convex combination of permutation matrices:

LEMMA 3.1. *The $n \times n$ matrix is doubly stochastic if and only if there exist nonnegative numbers t_π such that*

$$A = \sum_\pi t_\pi P_\pi \quad \textit{and} \quad \sum_\pi t_\pi = 1, \tag{3.1}$$

where the sums are over all $n!$ permutations $\pi = \pi(1) \ldots \pi(n)$ of $\{1, \ldots, n\}$.

Proof. It is clear that every matrix of the form (3.1) is doubly stochastic, so the problem is to show that every doubly stochastic matrix can be represented in terms of permutation matrices. Such a decomposition relies on an important lemma published in 1935 by Philip Hall [**8**] (see also Egerváry's work of 1931 [**4**]), which we can formulate as follows:

LEMMA 3.2. *Consider n men and n women such that each man-woman pair is either 'compatible' or 'incompatible'. If there is no way to match the men and women into n compatible marriages, then for some $k > 0$ there is a set of k men who are compatible with only $k-1$ women.*

Proof. Suppose there is a way to obtain m compatible marriages but no way to obtain $m+1$, for some $m < n$, and let x be an unmarried man in one of these maximum matchings. Consider all chains of relationships of the form

$$i_1 \to j_1 \Rightarrow i_2 \to j_2 \Rightarrow \cdots \Rightarrow i_r \to j_r \tag{3.2}$$

where $x = i_1$, and where the relation $i \to j$ means 'man i is compatible with woman j' while $j \Rightarrow i$ means 'woman j is married to man i'. In every such chain, the woman j_r must be married; for if she were not, there would be a way to have $m+1$ compatible marriages, by performing $r-1$ divorces and then marrying i_l with j_l for $1 \leqslant l \leqslant r$. Consider now the set S of all men i_l appearing in chains (3.2), and the set T of all women j_l that appear. Then each woman in T is married to a man in S, and each man in S (except for x) is married to a woman in T. Therefore S contains k elements while T contains only $k-1$. ■

Returning now to the proof of Lemma 3.1, let A be doubly stochastic and let us imagine n men and women such that man i is compatible with woman j if and only if $a_{ij} > 0$. In these circumstances a set of n compatible marriages is possible; for if S were a set of k men that are compatible with only $k-1$ women, the sum $\Sigma\{a_{ij} \mid i \in S, j \in T\}$ would equal k because it includes all the nonzero elements of k rows, yet it would involve only $k-1$ columns so that it could not exceed $k-1$. Thus there is a permutation $\pi(1) \ldots \pi(n)$ such that $a_{i\pi(i)} > 0$ for $1 \leqslant i \leqslant n$. Let $t_\pi = \min(a_{1\pi(1)}, \ldots, a_{n\pi(n)})$; if $t_\pi = 1$, A is a permutation matrix, and so it clearly has the form (3.1). Otherwise

$$A = t_\pi P_\pi + (1 - t_\pi)B$$

for some doubly stochastic matrix B, where B contains at least one more zero than A. By

induction on the number of nonzero entries, B can be expressed in the form (3.1), and this yields the desired representation of A. ■

A doubly stochastic matrix $A = (a_{ij})$ in which a_{ij} represents the probability of going from state j to state i is said to have an *equilibrium vector* $x = (x_1, \ldots, x_n)$ if $Ax = x$, i.e., if $\Sigma_j a_{ij} x_j = x_i$ for all i. The matrix A is called *decomposable* if the set $\{1, \ldots, n\}$ can be partitioned into nonempty subsets S and T such that $a_{ij} = 0$ whenever $i \in S$ and $j \in T$. These concepts are related to each other by the following simple observation:

LEMMA 3.3. *If $x = (x_1, \ldots, x_n)$ is an equilibrium vector for A having some components unequal, then A is decomposable.*

Proof. If $(x_1, \ldots, x_n)$ is an equilibrium vector, so is $(x_1 + c, \ldots, x_n + c)$; hence we can assume that all components of x are nonnegative and that some are zero. Let $S = \{i \mid x_i = 0\}$ and $T = \{i \mid x_i > 0\}$; then $\Sigma_j a_{ij} x_j = x_i$ implies that $a_{ij} = 0$ whenever $i \in S$ and $j \in T$. ■

The proof of Lemma 3.3 uses only the fact that A satisfies the 'singly stochastic' property $\Sigma_j a_{ij} = 1$ for $1 \leqslant i \leqslant n$. Doubly stochastic matrices satisfy a much stronger condition:

LEMMA 3.4. *If A is doubly stochastic and if S and T are sets of decomposability for A, then $a_{ij} = 0$ unless $i, j \in S$ or $i, j \in T$.*

Proof. We have $\Sigma_{i,j \in S} a_{ij} = \Sigma_{i \in S} \Sigma_{1 \leqslant j \leqslant n} a_{ij} = \Sigma_{i \in S} 1 = \Sigma_{j \in S} 1 = \Sigma_{j \in S} \Sigma_{1 \leqslant i \leqslant n} a_{ij} = \Sigma_{i,j \in S} a_{ij} + \Sigma_{i \in T, j \in S} a_{ij}$, so a_{ij} must be zero when $i \in T$ and $j \in S$. ■

4. Minimal Matrices. For purposes of this discussion we shall say that an $n \times n$ matrix $A = (a_{ij})$ is *minimal* if it is doubly stochastic and if it has the smallest permanent among all $n \times n$ doubly stochastic matrices. There is at least one minimal matrix (i.e., the smallest value is actually achieved), since the permanent is a continuous function of the matrix elements and since the set of doubly stochastic matrices is a closed and bounded subset of n^2-dimensional space. Lemma 3.1 implies that the permanent of A is at least $\Sigma_\pi t_\pi^n$ for some nonnegative numbers with $\Sigma_\pi t_\pi = 1$, so the minimal permanent cannot be zero.

According to van der Waerden's conjecture [**15**] there is only one minimal matrix of order n, namely the matrix with $a_{ij} = 1/n$ for all i and j. We are not ready to prove the conjecture yet, but we shall see that standard methods of minimization give us rather strict conditions that minimal matrices must satisfy.

From now on we shall use the notation A_{ij} to stand for the $(n-1) \times (n-1)$ matrix obtained from A by removing row i and column j. Clearly

$$\operatorname{per}(A) = \sum_j a_{ij} \operatorname{per}(A_{ij}) = \sum_i a_{ij} \operatorname{per}(A_{ij}); \tag{4.1}$$

in fact either of these equations can be used to provide an alternative definition of the permanent.

Another basic fact about 'permanents of minors' is

$$\operatorname{per}(A + \epsilon B) = \operatorname{per}(A) + \epsilon \sum_{i,j} b_{ij} \operatorname{per}(A_{ij}) + O(\epsilon^2), \tag{4.2}$$

where $O(\epsilon^2)$ refers to ϵ^2 times a polynomial in ϵ and the elements of A and B. If A is doubly stochastic let us say that B is a *valid modification* for A if the row sums and column sums of B are zero and if $b_{ij} \geqslant 0$ whenever $a_{ij} = 0$; then $A + \epsilon B$ is doubly stochastic for all sufficiently small $\epsilon > 0$.

LEMMA 4.1. *If A is a minimal matrix and if B is a valid modification for A, then*

$$\sum_{i,j} b_{ij} \operatorname{per}(A_{ij}) \geqslant 0. \tag{4.3}$$

Proof. This inequality follows immediately from identity (4.2) and the definition of minimality. ∎

LEMMA 4.2. *A minimal matrix is indecomposable.*

Proof. Suppose that A is doubly stochastic and that $a_{ij} > 0$ only when $i, j \in S$ or $i, j \in T$, where S and T are nonempty and $S \cup T = \{1, \ldots, n\}$. We know that $\operatorname{per}(A) > 0$, so there is a permutation $\pi(1) \ldots \pi(n)$ with $a_{i\pi(i)} > 0$ for all i. Let s and t be elements of S and T, respectively, and let $B = (b_{ij})$ be a matrix that is entirely zero except that $b_{s\pi(s)} = b_{t\pi(t)} = -1$ and $b_{s\pi(t)} = b_{t\pi(s)} = +1$. Since $\pi(s) \in S$ and $\pi(t) \in T$, the matrix B is a valid modification for A; therefore

$$\operatorname{per}(A_{s\pi(s)}) + \operatorname{per}(A_{t\pi(t)}) - \operatorname{per}(A_{s\pi(t)}) - \operatorname{per}(A_{t\pi(s)}) \leqslant 0$$

by Lemma 4.1. But this cannot happen, since $\operatorname{per}(A_{s\pi(s)})$ and $\operatorname{per}(A_{t\pi(t)})$ are positive, while $\operatorname{per}(A_{s\pi(t)})$ and $\operatorname{per}(A_{t\pi(s)})$ are zero. For example, $\operatorname{per}(A_{t\pi(s)})$ is zero because the matrix $(A_{t\pi(s)})$ has k rows corresponding to S in which all nonzero entries occur in only $k-1$ columns corresponding to $S - \{\pi(s)\}$. ∎

LEMMA 4.3. *If A is a minimal matrix, then* $\operatorname{per}(A_{ij}) > 0$ *for all i, j.*

Proof. If $\operatorname{per}(A_{ij}) = 0$, it has some set S of $k > 0$ rows in which all nonzero entries occur in some $k-1$ columns, by Lemma 3.2. Let $T = \{1, \ldots, n\} - S$; note that T is nonempty, since $i \in T$. We can now permute the columns of A to obtain a matrix A' in which all of the nonzero entries for the rows of S appear in the columns of S. Clearly A' is also a minimal matrix, and by Lemma 3.4 it is decomposable, contradicting Lemma 4.2. ∎

The next property of minimal matrices is the key to everything that follows; it was first proved by Marcus and Newman in 1959 [**10**].

THEOREM 4.4. *If A is a minimal matrix and $a_{ij} > 0$, then* $\operatorname{per}(A_{ij}) = \operatorname{per}(A)$.

Proof. The basic idea is to consider the set of all B that are valid modifications for A and such that $b_{ij} = 0$ when $a_{ij} = 0$. The other values of b_{ij} are unconstrained except for the fact that row sums and columns are zero; therefore the inequality (4.3) can be used with the technique of Lagrange multipliers to prove that there exist constants $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_n$ such that

$$\operatorname{per}(A_{ij}) = \lambda_i + \mu_j \qquad \text{if } a_{ij} > 0. \tag{4.4}$$

However, we shall consider a purely combinatorial proof, because it is based on elementary principles showing that a very simple class of matrices B is sufficient to prove (4.4).

By permuting the columns, we can assume without loss of generality that $a_{ii} > 0$ for all i. Let us write $i \to j$ if $a_{ij} > 0$; thus $i \to i$. Since A is indecomposable there is a "path"

$$1 = j_0 \to j_1 \to \cdots \to j_l = j \tag{4.5}$$

from 1 to j for all j. We say that j is at distance l if the shortest such path is of length l. For every $j > 1$ at distance l, let $p(j)$ be the smallest index at distance $l-1$ such that $p(j) \to j$. Then the path (4.5) is unique if we insist that $j_{k-1} = p(j_k)$ for $1 \leqslant k \leqslant l$. The arcs $\{p(j) \to j \mid j > 1\}$ may be regarded as an oriented tree emanating from point 1. We say that i is an ancestor of j (and we write $i \prec j$) if we have $i = p(j)$ or $i = p(p(j))$ or $\ldots$; the notation $i \preccurlyeq j$ means that $i = j$ or $i \prec j$.

Now we can establish (4.4). First we set $\lambda_1 = 0$ and $\mu_1 = \operatorname{per}(A_{11})$. Then for the indices $j \geqslant 2$ in increasing order of their distance from 1, we can define μ_j so that (4.4) holds when $i = p(j)$, since the value of $\lambda_{p(j)}$ was previously defined; after μ_j has been set, we immediately define λ_j so that (4.4) holds when $i = j$.

The construction above assigns values to $\lambda_1, \ldots, \lambda_n$ and $\mu_1, \ldots, \mu_n$; we still must prove (4.4) for the pairs $(\hat{i}, \hat{j})$ such that $a_{\hat{i}\hat{j}} > 0$ and $\hat{i} \neq \hat{j}$ and $\hat{i} \neq p(\hat{j})$. Consider the matrix B whose entries are all zero except that $b_{\hat{i}\hat{j}} = 1$, and that

$$b_{jj} = (j \prec \hat{j}) - (j \preccurlyeq \hat{i}), \quad \text{for } 1 \leqslant j \leqslant n;$$
$$b_{p(j)j} = (j \preccurlyeq \hat{i}) - (j \preccurlyeq \hat{j}), \quad \text{for } 1 \leqslant j \leqslant n. \tag{4.6}$$

The sum $\Sigma_i b_{ij} = (j = \hat{j}) + ((j \prec \hat{j}) - (j \preccurlyeq \hat{i})) + ((j \preccurlyeq \hat{i}) - (j \preccurlyeq \hat{j})) = 0$, and the sum $\Sigma_j b_{ij} = (i = \hat{i}) + ((i \prec \hat{j}) - (i \preccurlyeq \hat{i})) + ((i \prec \hat{i}) - (i \prec \hat{j})) = 0$; in the latter formula we use the fact that for fixed i there is a value of j such that $i = p(j)$ and $j \prec \hat{i}$ if and only if $i \preccurlyeq \hat{i}$. Therefore B is a valid modification for A, and so is $-B$. According to Lemma 4.1, the sum $\Sigma_{i,j} b_{ij}\operatorname{per}(A_{ij})$ must be zero. The sum actually turns out to be

$$\sum_{i,j} b_{ij}(\operatorname{per}(A_{ij}) - \lambda_i - \mu_j) = \operatorname{per}(A_{\hat{i}\hat{j}}) - \lambda_{\hat{i}} - \mu_{\hat{j}},$$

because $\lambda_i \Sigma_j b_{ij} = \mu_j \Sigma_i b_{ij} = 0$, and because (4.4) holds for all pairs (i, j) such that $b_{ij} \neq 0$ except possibly for the given pair $(\hat{i}, \hat{j})$. Thus (4.4) holds in general.

Now we are ready to complete the proof. We have $a_{ij}\operatorname{per}(A_{ij}) = a_{ij}(\lambda_i + \mu_j)$ for all i, j; hence by (4.1) we have

$$\operatorname{per}(A) = \lambda_i + \sum_j a_{ij}\mu_j = \mu_j + \sum_i a_{ij}\lambda_i$$

for all i, j. In matrix notation, $\lambda + A\mu = \mu + A^T\lambda = \operatorname{per}(A)e$, where e is a column vector of all 1's. Since $Ae = A^Te = e$, we have $A^T\lambda + A^TA\mu = \operatorname{per}(A)e$ and $A\mu + AA^T\lambda = \operatorname{per}(A)e$; hence

$$\mu = A^TA\mu \quad \text{and} \quad \lambda = AA^T\lambda.$$

Now A is indecomposable and has 1's on the diagonal; so the matrices A^TA and AA^T are even less decomposable. Consequently Lemma 3.3 implies that $\lambda_1 = \cdots = \lambda_n$ and $\mu_1 = \cdots = \mu_n$. We originally chose $\lambda_1 = 0$; hence $\lambda_1 = \cdots = \lambda_n = 0$ and $\mu_1 = \cdots = \mu_n = \operatorname{per}(A)$. ■

David London proved in 1971 [**9**] that something also can be said about $\operatorname{per}(A_{ij})$ when $a_{ij} = 0$.

LEMMA 4.5. *If A is a minimal matrix, then* $\operatorname{per}(A_{ij}) \geqslant \operatorname{per}(A)$ *for all i and j.*

Proof. (The following proof is based on an idea of Henryk Minc [**11**].) Because of Lemma 4.4, we need only consider (i, j) such that $a_{ij} = 0$; without loss of generality, assume that $i = j = 1$ and $a_{11} = 0$. By Lemma 4.3, $\operatorname{per}(A_{11}) > 0$; hence we can assume that $a_{jj} > 0$ for $2 \leqslant j \leqslant n$. Let $B = I - A$; all elements of B such that $a_{ij} = 0$ are nonnegative, and the row and column sums are zero, so that B is a valid modification of A.

By Lemma 4.1 and equation (4.1) we have

$$\begin{aligned} 0 \leqslant \sum_{i,j} b_{ij}\operatorname{per}(A_{ij}) &= \sum_j \operatorname{per}(A_{jj}) - \sum_{i,j} a_{ij}\operatorname{per}(A_{ij}) \\ &= \operatorname{per}(A_{11}) + (n-1)\operatorname{per}(A) - n\operatorname{per}(A) = \operatorname{per}(A_{11}) - \operatorname{per}(A), \end{aligned}$$

since $\operatorname{per}(A_{jj}) = \operatorname{per}(A)$ for $j > 1$ by Lemma 4.4. ■

5. Egorychev's Theorem. The results of Section 2 can now be combined with those of Section 4 to complete the analysis.

LEMMA 5.1. *If A is a minimal matrix, then* $\operatorname{per}(A_{ij}) = \operatorname{per}(A)$ *for all i and j.*

Proof. Without loss of generality, assume that $i = j = n$; assume further that $\operatorname{per}(A_{nn}) > \operatorname{per}(A)$ and $a_{(n-1)n} > 0$. Then $n > 1$, and Corollary 2.4 implies that

$$\operatorname{per}(A)^2 = \operatorname{per}(a_1, \ldots, a_n)^2 \geqslant \operatorname{per}(a_1, \ldots, a_{n-2}, a_{n-1}, a_{n-1})\operatorname{per}(a_1, \ldots, a_{n-2}, a_n, a_n).$$

But $\operatorname{per}(a_1, \ldots, a_{n-2}, a_{n-1}, a_{n-1}) = \Sigma_j a_{(n-1)j}\operatorname{per}(A_{nj}) > \Sigma_j a_{(n-1)j}\operatorname{per}(A) = \operatorname{per}(A)$, and $\operatorname{per}(a_1, \ldots, a_{n-2}, a_n, a_n) = \Sigma_j a_{nj}\operatorname{per}(A_{(n-1)j}) \geqslant \Sigma_j a_{nj}\operatorname{per}(A) = \operatorname{per}(A)$; so we have $\operatorname{per}(A)^2 > \operatorname{per}(A)^2$, a sort of contradiction. ■

LEMMA 5.2. *If A is a minimal matrix of order n, with $a_{ij} > 0$ for all i and j except possibly when $i = n$, then $a_{ij} = 1/n$ for all i and j.*

Proof. We have $\operatorname{per}(a_1, \ldots, a_{n-2}, a_n, a_n) = \Sigma_j a_{nj} \operatorname{per}(A) = \operatorname{per}(A)$ by (4.1) and Lemma 5.1, and similarly $\operatorname{per}(a_1, \ldots, a_{n-2}, a_{n-1}, a_{n-1}) = \operatorname{per}(A)$. Therefore equality holds in (2.6), and Theorem 2.3 implies that $a_n = \lambda a_{n-1}$ for some λ. Clearly $\lambda = 1$ since A is doubly stochastic; hence $a_n = a_{n-1}$. Similarly, all rows of A are equal. Therefore all columns of A consist of identical elements. Therefore all elements of A are equal to $1/n$. ■

The pieces of the puzzle are almost all in place, and we only need to dispense with the hypothesis that $a_{ij} > 0$.

THEOREM 5.3. *If A is a minimal matrix of order n, then $a_{ij} = 1/n$ for all i and j; hence*

$$\operatorname{per}(A) = n!/n^n. \tag{5.1}$$

Proof. Let B be obtained from A by replacing some row a_i by some other row a_k; then $\operatorname{per}(B) = \Sigma_j a_{kj} \operatorname{per}(A_{ij}) = \operatorname{per}(A)$. Similarly, let C be obtained from A by replacing the row a_k by a_i; again $\operatorname{per}(C) = \operatorname{per}(A)$. Of course, we do not know that B and C are doubly stochastic; but the matrix $D = \frac{1}{2}(B + C)$ certainly is doubly stochastic, and $\operatorname{per}(D) = \frac{1}{4}(\operatorname{per}(B) + 2\operatorname{per}(A) + \operatorname{per}(C))$, so D is a minimal matrix.

By a finite number of 'averaging' steps like the ones that formed D from A, we obtain a minimal matrix E having the same bottom row as A, but with $e_{ij} = 0$ only if $i = n$ or $a_{1j} = \cdots = a_{(n-1)j} = 0$. Since A is indecomposable, we cannot have $a_{1j} = \cdots = a_{(n-1)j} = 0$; hence E is a minimal matrix satisfying the condition of Lemma 5.2. It follows that its bottom row a_n is $(1/n, \ldots, 1/n)$. Similarly, all rows of A have this value. ■

The preparation of this report was supported in part by National Science Foundation grant MCS-77-23738 and by Office of Naval Research contract N00014-76-C-0330. Reproduction in whole or in part is permitted for any purpose of the United States government.

References

1. A. D. Aleksandrov, K teorii smeshannykh ob'ëmov vypuklykh tel, Mat. Sb. (new series), 3 (1938) 227–251.

2. Garrett Birkhoff, Tres observaciones sobre al algebra lineal, Univ. Nac. Tucumán Rev. Ser. A, 5 (1946) 147–151.

3. Herbert Busemann, Convex Surfaces, Interscience, New York, 1958, pp. 51–56.

4. E. Egerváry, Matrixok kombinatorius tulajfonságairól, Mat. Fiz. Lapok, 38 (1931) 16–28.

5. G. P. Egorychev, Reshenie problemy van-der-Wardena dli͡a permanentov, Inst. Fiziki im. L. V. Kirenskogo, USSR Acad. Sci., Siberian branch, preprint IFSO-13M, Krasnoi͡arsk, 1980.

6. ______, Novye formuly dli͡a permanenta, Dokl. Acad. Sci. USSR, 254 (1980) 784–787.

7. Werner Fenchel, Inégalités quadratiques entre les volumes mixtes des corps convexes, C. R. Acad. Sci. Paris, 203 (1936) 647–650.

8. Philip Hall, On representatives of subsets, J. London Math. Soc., 10 (1935) 26–30.

9. David London, Some notes on the van der Waerden conjecture, Linear Alg. Appl., 4 (1971) 155–160.

10. Marvin Marcus and Morris Newman, On the minimum of the permanent of a doubly stochastic matrix, Duke Math. J., 26 (1959) 61–72.

11. Henryk Minc, Doubly stochastic matrices with minimal permanents, Pacific J. Math., 58 (1975) 155–157.

12. Rolf Schneider, On A. D. Aleksandrov's inequalities for mixed discriminants, J. Math. Mech., 15 (1966) 285–290.

13. R. P. Stanley, Two combinatorial applications of the Aleksandrov-Fenchel inequalities, Dept. of Mathematics, M.I.T., 1980 (submitted for publication).

14. Bernard Tessier, Du théorème de l'indexe de Hodge aux inégalités isopérimétriques, C. R. Acad. Sci. Paris Sér. A, 288 (1979) 287–289.

15. B. L. van der Waerden, Problem, Jahresber. Deutsch. Math.-Verein., 25 (1926) 117.

16. J. H. van Lint, Notes on Egoritsjev's proof of the van der Waerden conjecture, Memorandum 1981-01, Dept. of Mathematics, Eindhoven University of Technology, Eindhoven, 1981.

CAUCHY'S THEOREM

GEOFFREY C. BERRESFORD
Department of Mathematics, C. W. Post Center, Long Island University, Greenvale, NY 11548

It is common practice to motivate Cauchy's Theorem by a simpler version (assuming continuous differentiability) before the general (or Cauchy-Goursat) case. In a tradition that began with Riemann, the simpler version is most often proved by Green's Theorem. There is, however, a simpler proof, based instead on reversing the order of differentiation and integration. It is far from new, dating back to Cauchy (1825) and perhaps even to Gauss (1811) (see **[1]** and **[2]**). The idea appears in Hurwitz-Courant [3] (also, as referees have pointed out, in [4] and [5]), but its periodic rediscovery (as illustrated by the "note added in proof" at the end of [6]) indicates that it does not have the currency it deserves. This note attempts to show the intuitive appeal of the idea and to give other "rediscovered" versions. The two very interesting articles [2] and [6] show that the idea is not limited to motivation but can indeed be used to obtain the general case.

CAUCHY'S THEOREM. *Let $f(z)$ be holomorphic in a simply connected domain D and let $c_0(t)$ and $c_1(t)$ be two continuous mappings of* $[0,1]$ *into D which have common endpoints*; *i.e.*, $c_0(0) = c_1(0)$ *and* $c_0(1) = c_1(1)$. *Then*

$$\int_{c_0} f(z)\,dz = \int_{c_1} f(z)\,dz. \tag{1}$$

To simplify the proof we will define holomorphic as continuously differentiable and assume further that $c_0(t)$ and $c_1(t)$ are continuously differentiable and can be smoothly deformed into each other within D with their endpoints remaining fixed. That is, we assume the existence of a twice continuously differentiable complex function $C(s,t)$ mapping $[0,1]\times[0,1]$ into D such that for all $0 \leqslant t \leqslant 1$

$$c_0(t) = C(0,t) \quad \text{and} \quad c_1(t) = C(1,t)$$

and, for all $0 \leqslant s \leqslant 1$, $C(s,0)$ and $C(s,1)$ remain constant.

Proof. Set

$$I(s) = \int_{C(s,t)} f(z)\,dz = \int_0^1 f(C(s,t))\frac{\partial C}{\partial t}\,dt.$$

Then (1) is equivalent to $I(0) = I(1)$, which will be proved if we can show that $I'(s) = 0$ for $0 < s < 1$.

$$\begin{aligned} I'(s) &= \int_0^1 f'(C(s,t))\frac{\partial C}{\partial s}\frac{\partial C}{\partial t} + f(C(s,t))\frac{\partial^2 C}{\partial s\,\partial t}\,dt \\ &= \int_0^1 \frac{\partial}{\partial t}\left[f(C(s,t))\frac{\partial C}{\partial s}\right]dt = f(C(s,t))\frac{\partial C}{\partial s}\bigg|_{t=0}^{t=1}, \end{aligned}$$

which is zero because $C(s,1)$ and $C(s,0)$ are constants. □

Note that the assumption that $C(s,t)$ is twice continuously differentiable is needlessly strong. We need require only that C_s, C_t, C_{st}, C_{ts} are continuous and that $C_{st} = C_{ts}$. Even these conditions need only hold piecewise, in the sense that there exists a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ such that the conditions above hold on each strip $[0,1]\times[t_{k-1}, t_k]$. For then

Geoffrey C. Berresford received his Ph.D. in 1974 under the direction of H. P. McKean, Jr., at the Courant Institute at N.Y.U. He has taught at the State University of New York at Purchase and is now Associate Professor of Mathematics at Long Island University's C. W. Post Center. His main interest is in stochastic processes, but he has published articles also on algorithms, differential equations, and logic.—*Editors*

$$I'(s) = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \frac{\partial}{\partial t}\left[f(C(s,t)) \frac{\partial C}{\partial s} \right] dt$$
$$= \sum_{k=1}^{n} f(C(s,t)) \frac{\partial C}{\partial s}\bigg|_{t_{k-1}}^{t_k} = f(C(s,t)) \frac{\partial C}{\partial s}\bigg|_0^1 = 0$$

just as before.

If the contour c_0 is closed, by taking the contour c_1 to be the single point $z_0 = c_0(0)$ we obtain immediately from (1) the closed contour form of Cauchy's theorem:

$$\int_{c_0} f(z)\,dz = 0.$$

Matters are particularly simple if the domain D is star-shaped with respect to z_0, for then the deformation $C(s, t)$ can be taken to be

$$C(s,t) = sz_0 + (1 - s)c_0(t).$$

Dropping the restriction that the endpoints $C(s,0)$ and $C(s,1)$ remain fixed leads to the following alternative proof of the closed contour case.

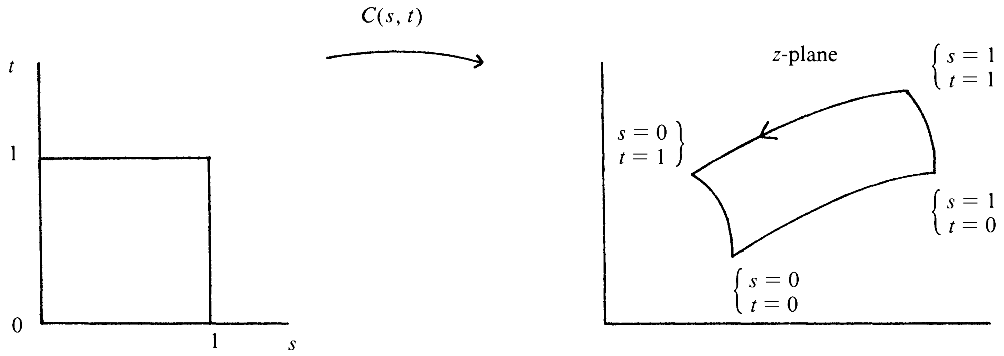
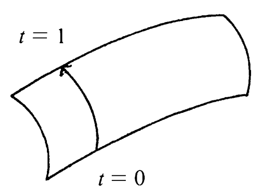


FIG. 1

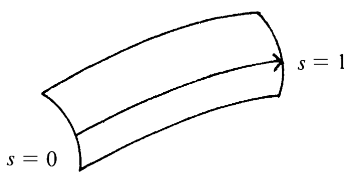
Let $C(s, t)$ be, as before, a twice continuously differentiable mapping of the square $[0, 1] \times [0, 1]$ into the z-plane (Fig. 1). For fixed s, integrating over $0 \leqslant t \leqslant 1$ gives

$$\int_0^1 f(C(s,t)) \frac{\partial C}{\partial t}\,dt, \qquad s \text{ fixed}$$



Similarly, for fixed t, integrating over $0 \leqslant s \leqslant 1$ gives

$$\int_0^1 f(C(s,t)) \frac{\partial C}{\partial s}\,ds, \qquad t \text{ fixed}$$



Then the integral over the entire closed contour C is

$$\begin{aligned}\int_C f(z)\,dz &= \int_{t=0}^{t=1} f(C(s,t))\frac{\partial C}{\partial t}\,dt\bigg|_{s=0}^{s=1} - \int_{s=0}^{s=1} f(C(s,t))\frac{\partial C}{\partial s}\,ds\bigg|_{t=0}^{t=1}\\ &= \int_{s=0}^{s=1}\frac{d}{ds}\int_{t=0}^{t=1} f(C)\frac{\partial C}{\partial t}\,dt\,ds - \int_{t=0}^{t=1}\frac{d}{dt}\int_{s=0}^{s=1} f(C)\frac{\partial C}{\partial s}\,ds\,dt\end{aligned}$$

and if as before we move the differentiations forward across the integral signs and differentiate the products, the double integrals cancel, proving $\int_C f = 0$. □

In this proof we may take the mapping $C(s, t)$ to be constant on one or more sides of the square. For example, in the mapping pictured below, $C(s, t)$ is constant on the bottom edge of the square (Fig. 2). The angle θ may be chosen arbitrarily (e.g., $C(s, t) = te^{-i\alpha s}$ gives an angle of α radians).

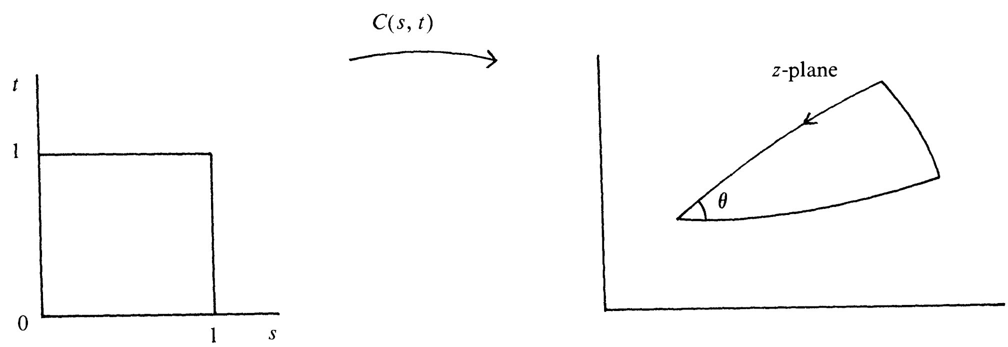


FIG. 2

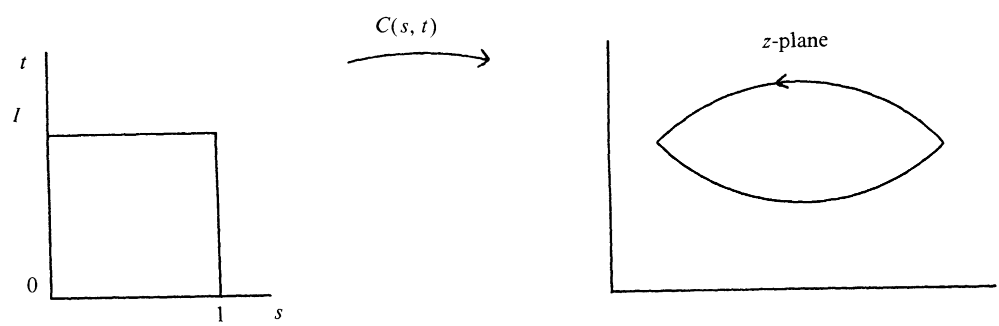


FIG. 3

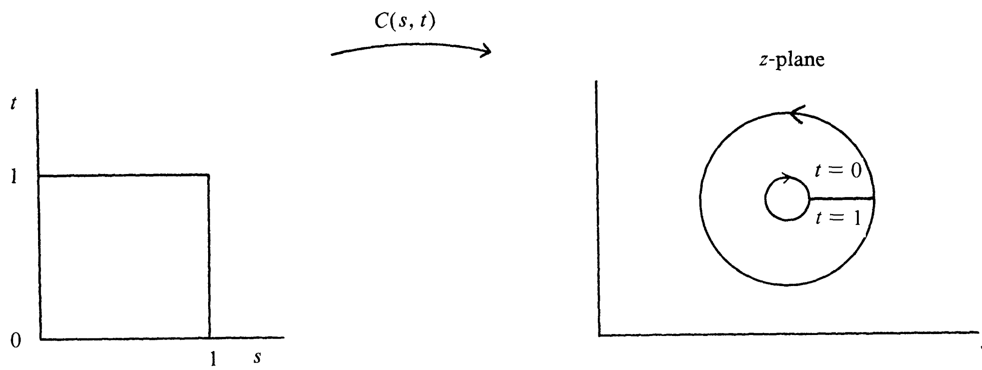


FIG. 4

In Fig. 3, $C(s, t)$ is taken constant on the top and bottom edges of the square. By enlarging the two angles to π radians we may obtain a circle (e.g., $C(s, t) = se^{i\pi t} + (1 - s)e^{-i\pi t}$).

Finally, if $C(s,0) = C(s,1)$ for $0 \leq s \leq 1$ we may obtain an annulus (Fig. 4) (e.g., $C(s, t) = (s + 1)e^{2\pi i t}$).

The author wishes to thank K. O. Friedrichs, who suggested many of these ideas, and the C. W. Post Research Committee, who provided financial assistance.

References

1. G. Birkhoff, ed., A Source Book in Classical Analysis, Harvard University Press, Cambridge, Mass., 1973.
2. W. F. Eberlein, It's All in the Family: Is Cauchy's Integral Theorem Trivial?, Rochester Mathematical Reports, #1, Rochester, N.Y., 1979.
3. Hurwitz-Courant, Functionentheorie, 3rd ed., Springer-Verlag, Berlin, 1929, pp. 284–285.
4. J. B. Conway, Functions of One Complex Variable, Springer-Verlag, New York, 1973, pp. 82–83.
5. E. B. Saff and A. D. Snider, Fundamentals of Complex Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1976.
6. L. Flatto and O. Shisha, A Proof of Cauchy's Integral Theorem, J. Approx. Theory, 7 (1973) 386–390.

THE HILBERT MODEL OF HYPERBOLIC GEOMETRY

HEINRICH GUGGENHEIMER

Department of Mathematics, Polytechnic Institute of New York, Brooklyn, NY 11201

The standard models of plane hyperbolic geometry are the Poincaré model in the upper half-plane or the unit circle and the Cayley-Klein model in the interior of a nondegenerate conic. The group of hyperbolic motions is a three-parameter group of complex fractional linear transformations for the Poincaré models and the projective group of three variables that leaves an indefinite quadratic form invariant for the Klein model. The isomorphism of the models (over a fixed real ordered field) is not difficult to establish: The two Poincaré models are isomorphic by a Möbius transformation, and the Poincaré model in the unit circle can be projected by a stereographic projection from the south pole into the upper hemisphere and then by an orthogonal projection back into the Klein model based on the unit circle. Since stereographic projection is nonlinear, the isomorphism of the groups has to be derived indirectly; a direct verification is possible but extremely laborious.

For axiomatic purposes, Hilbert ([**1**], [**2**]; and see [**3**], [**4**]) introduced a coordinate system based on the field of ends of hyperbolic geometry. The coordinate system yields a Klein model for which the group of hyperbolic motions is represented as that of the Poincaré model in the upper half-plane. This aspect has never been fully worked out, so it may be of interest to derive hyperbolic geometry in the Hilbert setting directly from the definition of the group and the isotropy group of a point.

Let F be a real ordered field in which every positive element has a square root. The group of hyperbolic transformations over F is the group of linear fractional transformations

$$w = \frac{az + b}{cz + d}, \quad a, b, c, d \in F, \quad ad - bc \neq 0. \tag{1}$$

The author received his doctorate under Beno Eckmann at Eidgenössiche Technische Hochschule Zürich (ETH). He has held positions at Hebrew University, Bar Ilan University (Israel), Washington State University, the University of Minnesota, and the Polytechnic Institute of New York (since 1967). His mathematical interests are in geometry of all kinds, with applications to topology, differential equations, etc.; and in history and philosophy of mathematics. His books include *Differential Geometry* (1963), *Plane Geometry and Its Groups* (1967), *Applicable Geometry* (1978), and *Mathematics for Engineering and Science* (1977).—*Editors*

The transformation (1) is given by the set of matrices

$$\lambda\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \lambda M, \qquad \lambda \in F, \;\; \lambda \neq 0.$$

A hyperbolic point reflection (or half-turn) is a transformation (1) of order two and positive determinant; a hyperbolic line reflection (or simply reflection) is a transformation (1) of order two and negative determinant. Now $M^2 = \lambda I$ if Tr $M = 0$; hence a point reflection can always be represented by

$$\lambda\begin{bmatrix} y & x \\ -1 & -y \end{bmatrix} \qquad x, y \in F$$

with

$$x > y^2. \tag{2}$$

That means that the points of our geometry can be uniquely given by the points $(x, y) \in F^2$ in the convex domain defined by the parabola $x = y^2$. Let F_* be the extension of F by the symbol ∞ with the usual rules. An end of the geometry is a couple $(y^2, y) \in F_*^2$.

A line reflection can be reduced to one of two normal forms,

$$\begin{bmatrix} -l_2 & -l_1 \\ 1 & +l_2 \end{bmatrix} \text{ or } \begin{bmatrix} -1 & -l_1 \\ 0 & +1 \end{bmatrix}.$$

As we shall see, the two normal forms correspond to lines that do not or do pass through the end (∞^2, ∞).

We identify a point P with the reflection in P; similarly a line l is identified with the reflection in l. The incidence relation $P \in l$ means $l(P) = P$ or, in terms of the reflections alone, $lPl^{-1} = P$; i.e., $lP = Pl$. In terms of the matrices this means

$$\begin{bmatrix} -l_2 & -l_1 \\ 1 & +l_2 \end{bmatrix}\begin{bmatrix} y & x \\ -1 & -y \end{bmatrix} = \lambda\begin{bmatrix} y & x \\ -1 & -y \end{bmatrix}\begin{bmatrix} -l_2 & -l_1 \\ 1 & +l_2 \end{bmatrix};$$

hence $\lambda = -1$ and the equation of a line of first kind is

$$x - 2l_2 y + l_1 = 0. \tag{3}$$

For the other kind of line we get from

$$\begin{bmatrix} -1 & -l_1 \\ 0 & +1 \end{bmatrix}\begin{bmatrix} y & x \\ -1 & -y \end{bmatrix} = \lambda\begin{bmatrix} y & x \\ -1 & -y \end{bmatrix}\begin{bmatrix} -1 & -l_1 \\ 0 & +1 \end{bmatrix}$$

that $\lambda = -1$ and

$$2y = l_1; \tag{4}$$

these are the parallels to the x-axis. Two coordinates (l_1, l_2) define a line (3) only if

$$l_2^2 > l_1; \tag{5}$$

that means that the points of F^2 outside the parabola $x = y^2$ represent lines. The line (3) intersects the parabola at the two ends given by

$$y_{1,2} = l_2 \pm \sqrt{l_2^2 - l_1}\,;$$

$y \in F$ by our hypotheses and

$$l_1 = y_1 y_2, \quad 2l_2 = y_1 + y_2.$$

The tangents to the parabola $x = y^2$ at (y_1^2, y_1) and (y_2^2, y_2) intersect at (l_1, l_2): *The coordinates of a line of first kind are the coordinates of the pole of the line relative to the parabola $x^2 = y$.* The

poles of the lines $y = \frac{1}{2} l_1$ are on the line at infinity of the projective plane over F. Two lines l and l' are perpendicular if the reflection in one is an automorphism of the other or $l(l') = l'$, i.e.,

$$\begin{bmatrix} -l_2 & -l_1 \\ 1 & l_2 \end{bmatrix}\begin{bmatrix} -l'_2 & -l'_1 \\ 1 & l'_2 \end{bmatrix} = -\begin{bmatrix} -l'_2 & -l'_1 \\ 1 & l'_2 \end{bmatrix}\begin{bmatrix} -l_2 & -l_2 \\ 1 & l_2 \end{bmatrix}$$

or

$$l_1 + l'_1 - 2l_2l'_2 = 0. \tag{6}$$

For the orthogonality of lines of first and second kind we get

$$\begin{bmatrix} -l_2 & -l_1 \\ 1 & l_2 \end{bmatrix}\begin{bmatrix} -1 & -l'_1 \\ 0 & 1 \end{bmatrix} = -\begin{bmatrix} -1 & -l'_1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -l_2 & -l_1 \\ 1 & l_2 \end{bmatrix}$$

$$l'_1 - 2l_2 = 0. \tag{7}$$

Two lines

$$x - 2l_2y + l_1 = 0$$

$$x - 2l'_2y + l'_1 = 0$$

intersect if $x > y^2$ for the point of intersection; i.e.,

$$4(l_1l'_2 - l_2l'_1) > \frac{(l_1 - l'_1)^2}{l_2 - l'_2}, \qquad l_2 \neq l'_2. \tag{8}$$

The lines are parallel for an $=$ sign in (8); they do not meet if

$$4(l_1l'_2 - l_2l'_2) < (l_1 - l'_1)^2(l_2 - l'_2)^{-1}.$$

A comparison of (3) and (6) shows that two lines are perpendicular in the Hilbert model if and only if one passes through the pole of the other.

For a geometry defined on an arbitrary real ordered field F with square roots one cannot define a metric in the usual sense since the logarithmic function is not defined on F. This is no disadvantage; we are led to develop hyperbolic trigonometry in a natural way. Length and angle are invariants of translations and notations.

Given two lines $l(l_1, l_2)$ and $l'(l'_1, l'_2)$, the product of the reflections in l and l' is (multiplication from right to left):

$$\lambda\begin{bmatrix} l_2l'_2 - l'_1 & l_1l'_2 - l'_1l_2 \\ l'_2 - l_2 & l_2l'_2 - l_1 \end{bmatrix}. \tag{9}$$

The action of a hyperbolic motion as a transform of a matrix (9) is by similarity. The only similarity invariants of a matrix M of rank 2 are $\operatorname{Tr} M$ and $\det M$. Therefore, every Möbius invariant of M is a function of

$$\frac{\operatorname{Tr} M}{2\sqrt{\det M}}$$

which is well defined since the determinant of every matrix (9) is > 0. A direct computation shows that (8) is equivalent with

$$\left|\frac{\operatorname{Tr} M}{2\sqrt{\det M}}\right| < 1. \tag{10}$$

A *rotation* is the product of two reflections in lines that intersect in the hyperbolic plane. A *parallel displacement* (or *horocyclic motion*) is the product of two reflections in lines that intersect at an end. A *translation* is the product of two reflections in lines that intersect in a point P outside the hyperbolic plane. Let p be the polar of P; put $Q = l \cap p,\ R = l' \cap p$. Let l denote also the

reflection in l; Q, that in Q. Then

$$l'l = l'ppl = (l'_p)(pl) = RQ.$$

If $Q=(x,y)$ and $R=(\xi,\eta)$, the translation can also be given by

$$\lambda\begin{bmatrix} \eta y-\xi & \eta x-\xi y \\ \eta-y & \eta y-x \end{bmatrix} = \lambda T. \tag{11}$$

A rotation defined by lines l, l' has the intersection $R = l \cap l'$ as invariant point. In euclidean geometry, the rotation invariant is the angle. The Hilbert model is not conformal but the rotation subgroup for the center $R(1,0)$ behaves like a euclidean rotation group. The matrix A belongs to that subgroup if

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \lambda\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

or $\lambda = 1$, $a = d$, $b = -c$. If we normalize the determinant to $+1$, the rotation matrix becomes

$$A = \begin{bmatrix} \cos A & \sin A \\ -\sin A & \cos A \end{bmatrix} \tag{12}$$

where

$$\cos A = \frac{\operatorname{Tr} A}{2\sqrt{\det A}}. \tag{13}$$

To see that this definition is adequate, we note that (6) and (9) imply $\operatorname{Tr} A = 0$ for $l \perp l'$. If $F = \mathbb{R}$, the rotations of center Q form a compact group which is similar to an orthogonal group (12). In the general case, the rotation group of center Q is similar to the group of matrices (12) by reflection in the midpoint $M(\eta, \xi)$ of $Q(x,y)$ and $R(I,0)$. Here ξ, η are determined by

$$\xi = \frac{x-1}{y}\eta + 1, \qquad \eta^2(y^2 - x + 1) = 2\eta y + y^2 = 0,$$

the equations are obtained from $M \in$ line RQ, $MQM^{-1} = R$. Since the addition formulas of trigonometry follow from matrix multiplication in the group (12), the formulas are valid in hyperbolic geometry with $\sin A = \sqrt{1-\cos^2 A}$. One has to note that the angles determined by the inverse trigonometric functions for $F = \mathbb{R}$ would refer to the angle between the lines, not the (euclidean) angle of rotation!

For a translation we have to define (for $F = \mathbb{R}$ by analytic continuation)

$$\cosh T = \frac{\operatorname{Tr} T}{2\sqrt{\det T}}. \tag{14}$$

The hyperbolic sine is defined by $\cosh^2 T - \sinh^2 T = 1$. To see that (14) is adequate, we note that three collinear points can be mapped into $P(a,0)$, $Q(b,0)$, $R(c,0)$. Then, by (11),

$$\begin{aligned} \cosh RP &= \frac{a+c}{2\sqrt{ac}} = \frac{a+b}{2\sqrt{ab}}\frac{b+c}{2\sqrt{bc}} + \frac{b-a}{2\sqrt{ab}}\frac{c-b}{2\sqrt{bc}} \\ &= \cosh QP \cosh RQ + \sinh QP \sinh RQ \\ &= \cosh(QP + RQ). \end{aligned} \tag{15}$$

Here again, the argument is not the (euclidean) translation vector but the ordered couple of defining points. A right triangle can be mapped onto $A(1,0)$, $B(a,0)$, $C(a,b)$. Then

$$\cosh CA = \frac{1+a}{2\sqrt{a-b^2}} = \cosh CB \cosh BA = \frac{\sqrt{a}}{\sqrt{a-b^2}}\frac{1+a}{2\sqrt{a}} \tag{16}$$

is the cosine theorem of hyperbolic geometry and

$$\sinh CB = \frac{b}{\sqrt{a - b^2}} = \sinh CA \cdot \sin\alpha = \frac{\sqrt{(a-1)^2 + 4b^2}}{2\sqrt{a - b^2}} \frac{2b}{\sqrt{(a-1)^2 + 4b^2}} \tag{17}$$

is the sine theorem. Here α is the rotation matrix defined by the line $y = 0$ and the line AC. A rotation of angle $\alpha/2$ transforms $y = 0$ into the line of matrix

$$\begin{bmatrix} \cos\frac{\alpha}{2} & \sin\frac{\alpha}{2} \\ -\sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\frac{\alpha}{2} & -\sin\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{bmatrix} = \sin\alpha \begin{bmatrix} -\cot\alpha & 1 \\ -1 & \cot\alpha \end{bmatrix}. \tag{18}$$

The equation of the line is, therefore,

$$x - 2y\cot\alpha - 1 = 0 \tag{19}$$

or

$$\tan\alpha = \frac{2y}{x-1};$$

in our case $\tan\alpha = 2b/(a-1)$ and $\sin\alpha = 2b((a-1)^2 + 4b^2)^{-\frac{1}{2}}$.

Finally, parallel displacements are defined by $\operatorname{Tr} M = 2\sqrt{\det M}$. Every parallel displacement is similar, in the Möbius group, to

$$\lambda \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Solid hyperbolic geometry can be developed along similar lines from the group of linear fractional transformations with coefficients in the complexified field F^c. The corresponding Hilbert model is linear in the paraboloid $x^2 + y^2 \leqslant z$. Another linear model in a paraboloid has been derived by D. Pedoe [**5**].

References

1. D. Hilbert, Neue Begründung der Bolyai-Lobatschefskyschen Geometrie, Math. Ann., 57 (1903) 137–150.

2. ______, Grundlagen der Geometrie, Anhang III, 10, Aufl., Teubner, Stuttgart, 1968.

3. F. Bachmann, Aufbau der Geometrie aus dem Spiegelungsbegriff, Springer, Berlin-Göttingen-Heidelberg, 1959, § 11.

4. H. Guggenheimer, Plane Geometry and Its Groups, Holden-Day, San Francisco (now Krieger, Huntington, N.Y.), 1967, Chap. 9.

5. D. Pedoe, Circles, Pergamon Press, New York, 1957.

THE IMPOSSIBILITY OF TILING A CONVEX REGION WITH UNEQUAL EQUILATERAL TRIANGLES

EDWIN BUCHMAN

Department of Mathematics, California State University, Fullerton, CA 92634

1. A square can be dissected into, or tiled by, a finite number of smaller squares in such a way that no two of the smaller squares are equal in size [**1**], [**3**]. Tutte [**5**] proved, as part of a broader investigation in graph theory, that an equilateral triangle cannot be tiled by smaller unequal equilateral triangles (although it can be tiled by smaller equilateral triangles, two of which have

---

The author received his Ph.D. from the University of California, Los Angeles, in 1968 under Frederick Valentine. Since then he has been on the mathematics faculty of California State University, Fullerton. His current research interests include convex sets, mathematical logic, and the theory of geometric constructibility.—*Editors*

the same size and orientation). (The trivial case of a single equilateral triangle "tiling" itself is excluded by the formal definition of a tiling.) Tutte's method can be extended to provide a simple proof that no convex region can be tiled by unequal equilateral triangles. This is done in Section 2. In Section 3 it is shown that any nonequilateral triangle can be tiled by smaller unequal triangles similar to itself.

2. By a tiling, or dissection, of a region $R \subset E^2$ we mean the expression of R as the union of a finite number (greater than one) of smaller regions which overlap only at their boundaries. In particular, we shall consider only tilings in which the smaller regions, or tiles, are triangles.

Consider a tiling of a convex region R by equilateral triangles. It is evident that the tiles fall into two classes according to their orientations, those of one class having orientation rotated through the angle π from the orientation of the triangles in the other class. The boundary of the region R must consist of segments parallel to the sides of the triangles in these orientations. Thus the boundary of R is a hexagon with interior angles equal to $2\pi/3$, except that one or more sides of the hexagon can degenerate to points where the interior angle is $\pi/3$.

If R is not already an equilateral triangle, we can transform R into an equilateral triangle T by adjoining equilateral triangles to three (or fewer) sides of R, as shown in Fig. 1. Note that T is tiled by these additional triangles and the tiles of R.

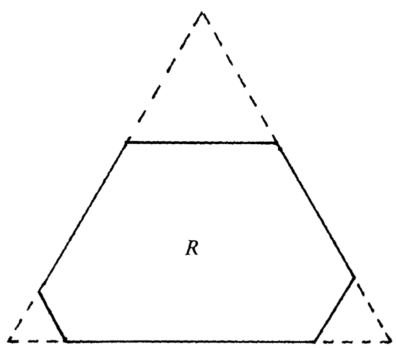


FIG. 1

Each tile of T occupies an angle $\pi/3$ at each of its vertices; so at each vertex v we have three possibilities: (1) v is a vertex of T, (2) six tiles have a vertex at v, or (3) exactly three tiles have a vertex at v, the remaining angle π being occupied by another tile or by the exterior of T. We form a planar graph Q in the following way. We place a red point at the center of each tile of T. We place a green point on each point where exactly three vertices of tiles lie. We do not place green points on the vertices of T, but we place one green point somewhere in the exterior of T, as shown in Fig. 2. We do not place a green point on any point where six tiles have a common vertex, but arbitrarily choose two of these six tiles, situated oppositely about their common vertex, and place one green point a small distance (compared with the sides of the tiles) away from this vertex into the interior of each of these two tiles, as shown in Fig. 3. The red and green points together form the nodes of Q.

The nodes are connected as follows to form the edges of Q: (1) we connect the red point in each tile to each of its vertices which bears a green point; (2) if a vertex of a tile is a vertex of T, we join the red point in that tile to the exterior green point by a path through the vertex of T and through the exterior of T; and (3) if a vertex of a tile is a point where six tiles meet, we join the red point in that tile to the green point near that vertex within that tile or within an immediately adjacent tile. (Cases (2) and (3) are illustrated in Figs. 2 and 3, respectively, where only the edges referred to are shown.) The edges thus formed should not intersect each other except at their endpoints.

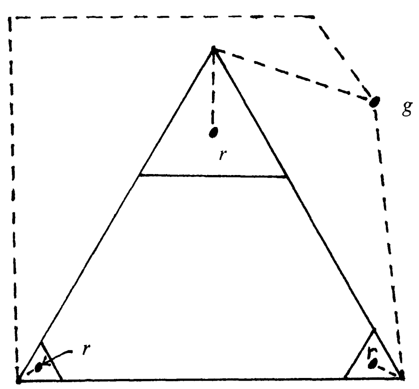


FIG. 2

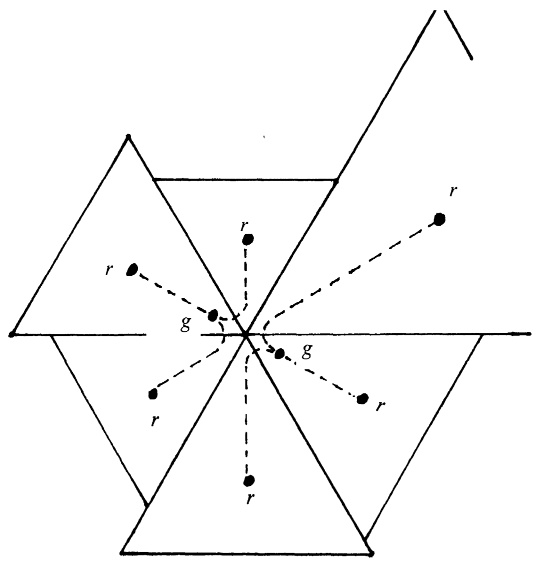


FIG. 3

Denote the number of nodes, edges, and faces of the graph Q respectively by V, E, and F. Let n be the number of tiles in T. Then there are n red nodes. Each edge connects a red node to a green node, and each red node has exactly three edges emanating from it; so $E = 3n$. Each green node also has exactly three edges in contact with it; so there must be n green nodes. Thus $V = n + n = 2n$. Euler's formula for planar graphs gives $F = E - V + 2 = n + 2$. Let c_i be the number of faces in Q which are bounded by exactly i edges. Since red and green nodes alternate around each face, $c_i = 0$ for odd values of i. Also, no pair of nodes is connected by more than one edge; so $c_2 = 0$. Thus

$$F = c_4 + c_6 + c_8 + c_{10} + \cdots = n + 2.$$

But each edge is counted twice by counting each edge in contact with each face; so

$$2E = 4c_4 + 6c_6 + 8c_8 + 10c_{10} + \cdots = 2(3n).$$

Eliminating n in the two preceding equations, we obtain

$$6 = c_4 - c_8 - 2c_{10} - \cdots;$$

so $c_4 \geqslant 6$—that is, there must be at least six four-sided faces in Q.

By an extreme tile we refer to any of the three tiles which have a vertex in common with T. The triangles, if any, added to R to produce T must be among the extreme tiles. By construction of Q, each four-sided face in Q corresponds to a separate instance of one of the following:

(a) a pair of extreme tiles having one vertex in common,

(b) a pair of tiles, one extreme and one not extreme, having two vertices in common,

or

(c) a pair of tiles, neither extreme, having two vertices in common.

Note that there can be at most three instances of (a) and three of (b). So either there is at least one instance of (c), in which case the two tiles must have an edge in common and both must come from the tiling of R, or there are exactly three instances each of (a) and (b), in which case it is easily seen that T is tiled by exactly four tiles of equal size, of which at least two must be from the tiling of R. In either of these cases, R must contain at least two tiles of equal size, proving the following:

THEOREM. *Any tiling of a planar convex region by equilateral triangles must have at least two triangles of equal size.*

Incidentally, a purely geometric proof of this theorem is not difficult but is more tedious than the graph theoretic proof above, and so it is only sketched here. Suppose that a convex region R is tiled by equilateral triangles, no two of which are equal. By a notch we mean a three-segment polygonal path which is a subset of the union of the boundaries of the tiles of R, whose angles are both $2\pi/3$ and whose two end segments lie on the same side of the line containing the middle segment. By the size of a notch we mean the length of its middle segment. It can be shown that there must be at least one notch; so there must be a smallest notch, $ABCD$ in Fig. 4. Repeatedly using the facts that $ABCD$ is the smallest notch and that the tiles are all unequal, it can be shown that BC must be the side of a single tile, which in turn must be bordered by exactly two tiles on each side, as illustrated. These, together with the shaded tiles, from three notches, $EFGH$, $IJKL$, and $MNOP$, whose average size equals the length of BC and which are therefore all equal in size. But this implies that tiles 1 and 2 are equal, a contradiction.

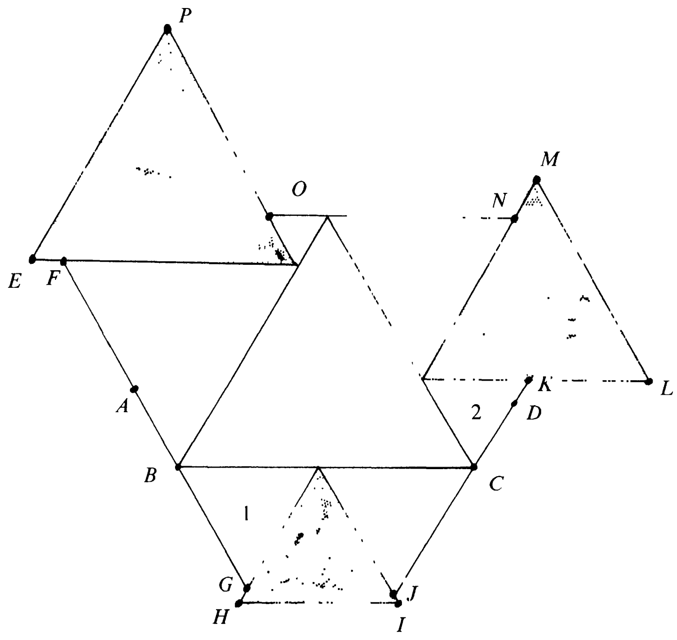


FIG. 4

3. *Nonequilateral triangles*. Every triangle other than an equilateral triangle has a tiling by unequal triangles similar to it. Such a tiling cannot have all of its tiles oriented the same as the large triangle or rotated through π from such orientation, because any similarity transformation of the plane which transforms the large triangle into an equilateral triangle would transform such a tiling into a tiling by unequal equilateral triangles. However, if there is no restriction on the orientation of the tiles, then any nonequilateral triangle has the tiling shown in Fig. 5. Here the vertices of the triangle are lettered so that side CB is longer than side AB. The angles of the smaller similar triangles are shown by a, b, and c. Triangle ADE of the tiling is oriented the same as triangle ABC. Using the length of AD as a unit, we denote the length of ED by $x > 1$. Other lengths can be computed as shown. The triangles of this tiling are all unequal except possibly for CFG and BDG, which will be unequal if $x + x^3 \neq x^4$. Thus the tiling of Fig. 5 can be used by suitably lettering the vertices of the triangle, unless the sides of the triangle are in the ratio $1 : 1 : r$

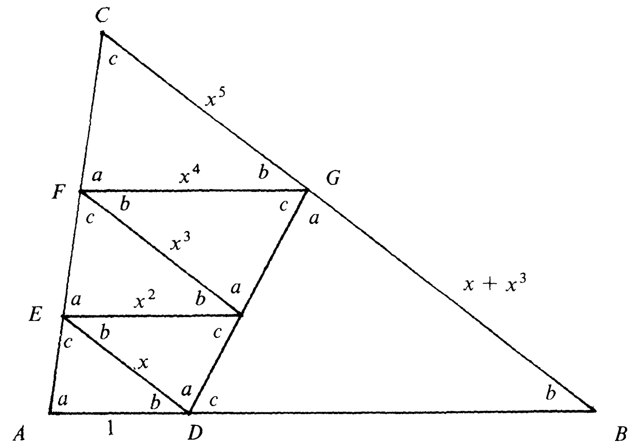


FIG. 5

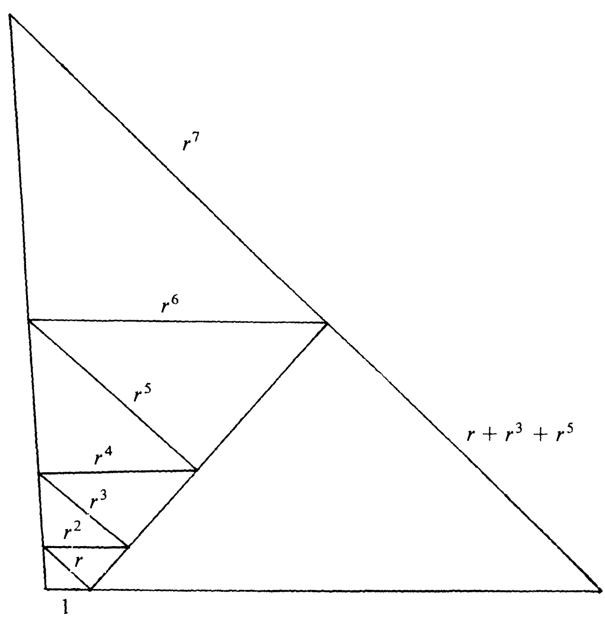


FIG. 6

or $1 : r : r$, where $r \approx 1.46557$ is the positive root of $x + x^3 = x^4$. In these two exceptional cases, a similar-unequal tiling can be constructed as shown in Fig. 6, drawn to illustrate the $1 : 1 : r$ case. Note that all of these tiles are unequal because $r > 1$, except for the largest two, whose comparable dimensions are r^6 and $r^5 + r^3 + r$, as shown. But these are unequal since $r > 0$ satisfies $x + x^3 = x^4$.

I would like to thank Professor Robert Sorgenfrey for the pleasant discussions that started me on this paper.

References

1. R. L. Brooks, C. A. B. Smith, A. H. Stone, and W. T. Tutte, The dissection of rectangles into squares, Duke Math. J. (1940) 312–340. (This paper also proves that a rectangular parallelepiped cannot be dissected into unequal

cubes, and it reports the result, later proved in [5], that an equilateral triangle cannot be tiled by unequal equilateral triangles.)

2. ______, Leaky electricity and triangulated triangles, Philips Res. Rep., 30 (1975) 205–219 (further discussion of the results of [5] and their relation to electric circuit theory).

3. Martin Gardner, Mathematical games, Scientific American (November 1958) 136–142 (an account by William Tutte of the "squaring the square" problem).

4. ______, Mathematical games, Scientific American (April 1965) 128–132 (the impossibility of "solidly tiling" a cube by unequal cubes).

5. W. T. Tutte, The dissection of equilateral triangles into equilateral triangles, Proc. Cambridge Phil. Soc., 44 (1948) 463–482.

A NOTE ON GOLOMB'S "CYCLOTOMIC POLYNOMIALS AND FACTORIZATION THEOREMS"

KATHERINE E. McLAIN AND HUGH M. EDGAR

Department of Mathematics, San Jose State University, San Jose, CA 95192

The object of this note is to show how a theorem of Guerrier [1] can be used to give a sharper form of Theorem 2 as found in Golomb [2]. Complete proofs can be found in McLain [3].

Let $n \geqslant 3$ be an integer and let q be a prime. We write $n = q^r N$ where $r \geqslant 0$ and $N \geqslant 1$ are integers with $q \nmid N$. Let F_q denote the finite field of q elements, ϕ the Euler totient function and finally $f = \operatorname{ord}_N q$ the order of q in the multiplicative group $(Z/NZ)^*$. Then we have

GUERRIER'S THEOREM. *The nth cyclotomic polynomial $\Phi_n(X)$ factors as follows with respect to $F_q[X]$:*

$$\Phi_n(X) = \Phi_{q^r N}(X) = [\Phi_N(X)]^{\Phi(q^r)} \tag{1}$$

where, further, $\Phi_N(X)$ factors as a product of $\phi(N)/f$ distinct irreducible factors, each of degree f.

If r is positive then $\Phi_n(X)$ is reducible with respect to $F_q[X]$ unless $(r, q) = (1,2)$ in which case $\Phi_n(X) = \Phi_N(X)$. However, for fixed n there are only finitely many primes q for which r is positive.

With the help of Dirichlet's theorem on primes in arithmetic progression we can now formulate our sharper form of Golomb's Theorem 2.

THEOREM 2′. *Assume that $q \nmid n$. If n has no primitive roots then $\Phi_n(X)$ is reducible with respect to $F_q[X]$ for all primes q. Otherwise, if $n = 4$, p^k or $2p^k$ for some odd prime p and some positive integer k then $\Phi_n(X)$ is irreducible with respect to $F_q[X]$ if and only if $q \equiv g \pmod n$ for some primitive root g of n. The (asymptotic) density of this set of primes—relativized with respect to the set of all primes—is given by $\phi(\phi(n))/\phi(n)$. In particular, $\Phi_n(X)$ is irreducible with respect to $F_q[X]$ for infinitely many primes q while $\Phi_n(X)$ is also reducible with respect to $F_q[X]$ for another infinite set of primes q.*

References

1. W. J. Guerrier, The factorization of the cyclotomic polynomials mod p, this MONTHLY, 75 (1968) 46.

2. Solomon W. Golomb, Cyclotomic polynomials and factorization theorems, this MONTHLY, 85 (1978) 734–737.

3. Katherine McLain, Irreducibility criteria, master's thesis, San Jose State University, 1980.

ON FUNCTIONS THAT ARE MONOTONE ON NO INTERVAL

TIBOR ŠALÁT

816 31 Bratislava, Mlynská Dolina, Czechoslovakia

Let $a, b \in R$. Denote by $C(a, b)$ the linear normed space of all continuous functions $f:[a, b] \to R$ with the norm $\| f \| = \sup_{a \leqslant t \leqslant b} |f(t)|$. Denote by $D(a, b)$ the linear normed space of all functions f: $[a, b] \to R$ with bounded derivative on $[a, b]$ (we put $f'(a) = f'_+(a), f'(b) = f'_-(b)$) with the norm

$$\| f \| = \sup_{a \leqslant t \leqslant b} |f(t)| + \sup_{a \leqslant t \leqslant b} |f'(t)|. \tag{1}$$

It is well known that each of the spaces $C(a, b)$, $D(a, b)$ is a Banach space ([**2**], p. 84; [**7**]).

In the paper [**3**] there is a construction of an example of such a function, with bounded derivative, that is monotone on no interval. In connection with this construction the following question arises: Denote by $C_0(a, b)$ and $D_0(a, b)$ the class of all functions $f \in C(a, b)$ and $f \in D(a, b)$, respectively, that are monotone on no interval. How large are the sets $C_0(a, b)$ and $D_0(a, b)$ in the respective spaces $C(a, b)$ and $D(a, b)$?

The following theorems give the answer to this question.

THEOREM 1. *The class $C_0(a, b)$ is a residual set in the space $C(a, b)$.*

THEOREM 2. *The class $D_0(a, b)$ is a nowhere dense set in the space $D(a, b)$.*

The comparison of Theorems 1 and 2 shows that the class of all functions $f \in C(a, b)$ that are monotone on no interval is a rich set in $C(a, b)$ (from the topological point of view) while the class of all such functions in the space $D(a, b)$ is a poor set (in $D(a, b)$).

An elementary proof of Theorem 1 can be found in [**5**, (pp. 296–298)]. Another proof of Theorem 1 can be obtained also by help of the well-known Banach theorem ([**2**, p. 260]) using the Lebesgue theorem on derivatives of monotone functions ([**2**, p. 264])—see [**5**, p. 298].

We shall give a proof of Theorem 2. We recall that a function $f \in D(a, b)$ is said to be a function of type P_1 (on $[a, b]$) if the set $Z(f') = \{x \in [a, b]; f'(x) = 0\}$ is dense in $[a, b]$ (cf. [**7**]). Denote by $P_1(a, b)$ the class of all functions $f \in D(a, b)$ that are of the type P_1 (on $[a, b]$).

Proof of Theorem 2. Let $f \in D_0(a, b)$. Then it can easily be verified that $Z(f')$ is a dense set in $[a, b]$ (see the proof of Theorem 3 in [**7**, p. 4]). Indeed, f' changes its sign on each interval and f' has the Darboux property. Hence we have

$$D_0(a, b) \subset P_1(a, b). \tag{2}$$

Using the procedure from the proof of Theorem 14 of [**7**, pp. 21–22] we shall show that $P_1(a, b)$ is a closed set in $D(a, b)$.

Let $f_n \in P_1(a, b)$ $(n = 1, 2, \dots)$, $f \in D(a, b)$, and $\| f_n - f \| \to 0$ $(n \to \infty)$. It follows from the definition (1) of the norm in $D(a, b)$ that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f and $\{f'_n\}_{n=1}^{\infty}$ converges uniformly to f'. Since the derivatives f'_n $(n = 1, 2, \dots)$ belong to the first Baire class, each of the sets $Z(f'_n)$ $(n = 1, 2, \dots)$ is a G_δ-set in $[a, b]$ and, since $f_n \in P_1(a, b)$ $(n = 1, 2, \dots)$, each of these sets is dense in $[a, b]$. Hence $Z(f'_n)$ $(n = 1, 2, \dots)$ is a residual set in $[a, b]$ (cf. [**6**, p. 49]). But then the set $\cap_{n=1}^{\infty} Z(f'_n)$ is also a residual set and is clearly a subset of the set $Z(f')$. Hence $Z(f')$ is dense in $[a, b]$ and so $f \in P_1(a, b)$.

If $f, g \in P_1(a, b)$, then according to the previous considerations the set $Z(f') \cap Z(g')$ is a

residual set in $[a, b]$. Hence this set is dense in $[a, b]$ and clearly it is a subset of the set $Z(f' + g')$. From this it follows easily that $f + g \in P_1(a, b)$. Further, if $f \in P_1(a, b)$, then evidently $cf \in P_1(a, b)$ for each $c \in R$. Hence $P_1(a, b)$ is a closed linear subspace of the space $D(a, b)$, different from $D(a, b)$.

Now, if E is an arbitrary linear normed space and if $F \subset E$ is a closed linear subspace of E different from E, then F is a nowhere dense set in E (cf. [**1**]; [**4**, p. 37, Exercise 4]). Therefore $P_1(a, b)$ is a nowhere dense set in $D(a, b)$. The assertion follows from (2).

References

1. C. Goffman, Functions of finite Baire type, this MONTHLY, 67 (1960) 164–165.

2. E. Hewitt and K. Stromberg, Real and Abstract Analysis, Springer-Verlag, Berlin-Heidelberg-New York, 1969.

3. Y. Katznelson and K. Stromberg, Everywhere differentiable nowhere monotone functions, this MONTHLY, 81 (1974) 349–354.

4. W. Kolodziej, Selected Chapters from Mathematical Analysis (Polish), PWN, Warszawa, 1970.

5. W. Köhnen, Einführung in die Theorie der metrischen Räume, Aloys Henn Verlag, Kastellaun/Hunsrück, 1978.

6. C. Kuratowski, Topologie I, PWN, Warszawa, 1948.

7. S. Marcus, Sur les dérivées dont les zéros forment un ensemble frontière partout dense, Rend. Circ. Mat. Palermo, 12 (1963) 1–36.

UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

In this department the MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

RESEARCH PROBLEMS BECOME UNSOLVED PROBLEMS

RICHARD K. GUY

Department of Mathematics & Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4

What's in a name? As I prepare to serve under a fourth MONTHLY editor it seems a good moment to take stock. Two of the four wanted to phase out the section. The obvious objection to it under its old name was that the MONTHLY is not a research journal. On the other hand, unsolved problems, particularly those that are explicable to the mathematical man in the street, are the life-blood of mathematics. In [1973, 1120] (numbers in brackets are year and page numbers of MONTHLY articles) I wrote "... it is important for students and teachers to realize that though they may not (yet) feel capable of mathematical research, there are plenty of unsolved problems that are well within their comprehension. Many amateurs are attracted to the subject, and many successful researchers first gained their confidence by examining intuitive problems in areas such as euclidean geometry, theory of numbers and, latterly, combinatorics and graph theory, where it is possible to understand questions (even to formulate them) and to obtain original results without a deep prior theoretical knowledge." To avoid being repetitive about policy, including that concerning solutions, we ask readers to look at earlier articles [1971, 1113; 1973, 1120; 1975, 995; 1977, 807; and 1979, 847].

Which reminds me that our new editor tells me that he has an aversion to *lists*, and although he was polite about it I suspect that he includes these articles in that category. I sympathize, and have tried to adopt a more listless style this time. On the other hand, many correspondents and people I meet tell me that they find these articles to be a very useful and economical pooling of effort and information. I believe it was Claude Shannon who claimed that he would rather spend two or three days rediscovering a theorem than an hour or two searching it out in a library. Very praiseworthy, but not all of us have the time or the expertise.

This seems to be a good place to announce the appearance of collections of unsolved problems, and a notable one is the recent book of Erdős and Graham (1980). Dates or "tbp" (to be published) in parentheses refer to the list (!) at the end of this article. Springer-Verlag is launching a new series of problem books: the first of these, by the present writer (1981), should be available some months before the appearance of this article. For those who do not wish to read further, see Hammer (1977); the rest of us will enter the lists (but only in joust).

Richard Duke [1969, 1128] enunciated the now notorious Ringel-Kotzig conjecture that every tree has a β**-labelling** (or **graceful numbering**, as Golomb later dubbed it). That it is possible to label the n vertices of a tree with $1, 2, \ldots, n$ so that the absolute differences of the labels at the ends of each edge are exactly the numbers $1, 2, \ldots, n-1$. I feel like echoing the old newspaper editors who said "this correspondence must now cease." Huang, Kotzig, and Rosa (tbp) summarize the state of affairs very well.

> This conjecture remains unresolved to this day and has acquired the status of another "graphical disease." As is the case with serious diseases, an instant "cure" is rarely forthcoming, in spite of massive attempts: the authors know of some seventy papers devoted to one or other aspect of this disease; references to most of these can be found in a recent survey by Bermond (1979). However, many of these merely rediscover some early results.
>
> The present paper cannot offer a definite cure to this disease, either. It is, however, our firm belief that the results contained in this paper are not only new but mean a step in the right direction toward an ultimate cure of the labelling disease. Some of the results have been announced earlier but proofs have not been published before.
>
> Our leitmotif is as follows: although all trees are conjectured to have a β-labelling, not all trees have an α-labelling. On the other hand, in a certain sense, almost all trees have a α-labelling (Kotzig, 1973). In our opinion, an inductive proof of the existence of a β-labelling for every tree would somehow have to involve combining trees with α-labellings with trees admitting only β-labellings. Thus it seems important to exhibit classes of trees without α-labellings (we exhibit several such classes) and showing that they nevertheless admit a β-labelling. We feel that this is the main contribution of this paper.

Here, an α**-labelling (strongly graceful numbering)** is a β-labelling for which there is a fixed real number which lies between the labels at the ends of every edge. The only present contribution is to replace the lower bound of 70, in the first quoted paragraph, by 100.

The Genocchi numbers have $2t/(e^t + 1)$ for their exponential generating function. Gandhi [1970, 505] conjectured that they satisfy a recurrence involving the difference of squares, and several people [1971, 1117] confirmed this. The connection with another combinatorial problem is discussed by Françon and Viennot (1979). Here is a translation into English of the English translation of their abstract: "The type (set of values of peaks, troughs, double rises and falls) of a permutation on n letters is characterized by a map γ. The number of possible types is the Catalan number. The number of permutations whose type is associated with γ is $\gamma(1)\gamma(2)\ldots\gamma(n)$. Various classes of permutations provide enumeration formulas for the classical combinatorial numbers, e.g., those of Euler and Genocchi." See also the paper of Strehl (1979) where Gandhi's name is misspelt throughout.

Joseph Hammer's paper (1979) on lattice points and area-diameter relations, related to Wills's problem [1971, 47], reminds me to recommend Hammer's book (1977), an excellent source of problems on lattice points.

Dekking's paper [1979, 848 & 850] has now appeared in a more accessible form (1979). He answers some of Brown's questions [1971, 886] by showing that there is a sequence on 2 symbols

in which no 4 blocks occur consecutively which are permutations of each other and that there is a sequence on 3 symbols in which no 3 blocks occur consecutively which are permutations of each other. The corresponding 4,2 problem remains open. See also the paper of Prodinger and Urbanek (1979).

What is known about the footballers of Croam [1972, 1018], and matches between larger odd numbers of contestants is reviewed in Norman Biggs's (1979) paper. The graph O_k has for vertices the set of $(k-1)$-subsets of a set of cardinality $2k-1$, and two vertices are adjacent just if the subsets are disjoint. Biggs asks a dozen questions. If G is a distance-transitive graph and its intersection array is the same as that of O_k, is $G = O_k$? If G is distance-transitive and has the same spectrum as O_k, is $G = O_k$? Is O_k a Cayley graph for any $k \geqslant 4$? For which graphs G does a symmetric group S_n act on the vertices as a distance-transitive group of automorphisms? If G is distance-transitive with valence $k \geqslant 4$ and Aut G acts primitively on the vertices, is the diameter of G at most $k-1$, and is the bound attained only when $G = O_k$? For which classes of symmetric circuits on O_k is there a regular map with these circuits as faces? If $k \geqslant 4$, does O_k have $\lfloor k/2 \rfloor$ disjoint Hamilton circuits? If k is neither 3 nor a power of 2, is O_k edge k-colorable? What is the chromatic polynomial of O_k? Is there a perfect 1-code in O_k for any k other than 4 or 6? What is the minimum size of a maximal independent set in O_k? What is the maximum size of a 1-code in O_k?

Erdős and Guy [1973, 52] asked for the smallest crossing number, $g(n, m)$, of a graph with n vertices and m edges and conjectured that

$$c_1 m^3/n^2 < g(n, m) < c_2 m^3/n^2$$

Newborn and Moser (1980) proved that the largest number $H(n)$, of crossing-free Hamilton circuits in any rectilinear drawing of the complete graph, K_n, is at least $3(10^{\lfloor n/3 \rfloor})/20$ and conjectured that $H(n)$ grows only exponentially with n. Ajtai, Chvátal, Newborn, and Szemerédi (tbp) showed that

$$g(n, m) \geqslant m^3(n-2)(n-3)/64n^2(n-1)^2$$

and deduced the truth of the Newborn-Moser conjecture.

MacMahon's "prime numbers of measurement" are $m_1 = 1$ and m_n, the least positive integer not the sum of one or more consecutive terms of $m_1, m_2, \ldots, m_{n-1}$. If $A = \{a_1 < a_2 < \cdots\}$ is a sequence of positive integers such that every positive integer can be represented as the sum of one or more consecutive elements of A, then Porubský (1977) shows that for every $\alpha > 1$,

$$\liminf a_n \ln\ln n / \left(n(\ln n)^\alpha\right) = 0.$$

This is to be compared with George Andrews's conjecture [1975, 922] that $\lim m_n \ln\ln n/(n \ln n) = 1$.

Simeon Reich writes to say that his problem [1976, 266; 1980, 292] has been solved negatively by Dale Alspach (tbp), who shows that the weakly compact convex set $\{f \in L^1[0, 1]: \int_0^1 f(x)\,dx = 1, 0 \leqslant f(x) \leqslant 2 \text{ a.e.}\}$ does not have the fixed point property for nonexpansive mappings. It remains an open question whether there is a weakly compact convex subset of a reflexive space that lacks this property.

For a sequel to Mycielski's articles [1977, 723; 1978, 263] asking if one can solve equations in groups, see Roger Lyndon's paper (1980), which has an extensive bibliography. And in connexion with Phyllis Cassidy's Classroom Note [1979, 772], Lyndon notes that Gordon, Guralnick, and Miller (1978) state that there is a group of order 96 in which not every element of the commutator subgroup is a commutator and that 96 is the smallest such order.

Kotzig and Laufer [1978, 364] call a $(2k+1)$-vector, $\mathbf{p}$, a **permutation** if its coordinates are those of the **fundamental permutation** $\mathbf{f} = (-k, -k+1, \ldots, k-1, k)$ in some order, and call $\mathbf{p}$ a **σ-permutation** if $\mathbf{p} + \mathbf{f}$ is also a permutation. They also call two σ-permutations, $\mathbf{u}$, $\mathbf{v}$, an **additive pair** if $\mathbf{u} + \mathbf{v} + \mathbf{f}$ is a permutation. The values of $2k+1$ for which additive pairs exist include all

numbers 5^a7^b. Jean Turgeon (1980) extends this to $5^a7^b13^c$ and D. G. Rogers (tbp) announces that F. W. Roush's computer search enables this to be further extended to $5^a7^b13^c19^d$.

In answer to a question of Carl McCarty [1978, 578], Garner and Herzberg (1981) give a procedure for constructing latin queen squares for any prime order greater than or equal to 11.

Affirmative answers to Raphael Robinson's question [1979, 690] (Is the smallest order of any B_i in a solution of $(B_1 - 1)(B_2 - 1) \cdots (B_n - 1) = 0$, where the B_i are unknowns chosen from a finite abelian group (written multiplicatively), less than n?) were received, in chronological order, from Alfred W. Hales; Ki Hang Kim & Fred W. Roush; Sidney C. Garrison, Martin R. Pettet & Stephen M. Gagola; Masao Kiyota & Kazumasa Nomura; Geoffrey R. Robinson; and Gennady Lyubeznik. Raphael Robinson adds that the problem he stated was intended to be just an example of an unsolved problem about the equation. He hoped that someone would be able to find a more or less complete theory of the equation. He asks another question about the least order of any B_i. When $n = p + 1$ and p is prime, there is a solution in which each B_i has order p (Robinson, 1979, p. 241) so that the bound $n - 1$ is best possible. Can a better bound be obtained in other cases? Perhaps the answer is the largest prime less than n?

Alter [1980, 43] asked, in effect, for a completion of the information about congruent numbers up to 1000, and a proof that if $a \equiv 5$, 6, or 7 (mod 8), then a is congruent. An integer a is **congruent** if there are nontrivial solutions of the simultaneous diophantine equations

$$x^2 + ay^2 = z^2, \quad x^2 - ay^2 = t^2.$$

It is sufficient to look at squarefree values of a. Knowledge of congruent numbers is widely scattered in time and space; it is hard to be complete and up to date. Some results have been obtained constructively (the numbers involved are often astronomical) and are not always published in easily accessible places. There are several omissions and a few errors in Alter's paper (457 appears in both Table 1 and Table 2, for example) and a number of further results have come to light. Jean Lagrange (1976, 1977) is an indefatigable worker on this problem, and he continues to add new results. J. A. H. (*Fun with Figures*) Hunter has obtained or stimulated several constructions. I am indebted to them and to others for the information compiled in Table 1, which is reproduced from section D27 of the present writer's book (1981) by permission of the publisher, Springer-Verlag. A symbol in column c and row r refers to the number $a = 40c + r$. Here is a guide to the symbols.

□	a is not squarefree
C	a is congruent
N	a is noncongruent

C5, C7: primes $p \equiv 5$ or 7 (mod 8) are known to be congruent. N3, N2, N9, NX, NL: numbers of the following forms, where p and q are primes, are known to be noncongruent: $p \equiv 3 \pmod 8$; $2p$ with $p \equiv 9 \pmod{16}$; pq with $p \equiv q \equiv 3 \pmod 8$; $2p$ with $p \equiv 5 \pmod 8$; $2pq$ with $p \equiv q \equiv 5 \pmod 8$.

A, G, B, &, and J refer to sources: a thousand-year-old Arab manuscript; Gérardin; Bastien; Alter et al.; and Jean Lagrange.

A blank (or 1 in case $a \equiv 1$ (mod 8) is prime) means that the status of a is unknown. There remain only 7, 6, 2, 7, 0, and 1 such entries corresponding to $a \equiv 1, 2, 3, 5, 6$, and 7(mod 8), a total of 23. There are at most 8 numbers for which the conjecture is in doubt.

Ronald Alter points out misprints in his paper [1980, 206] with Jeffrey Barnett on the postage stamp (extremal basis) problem. On line 9 from the foot of p. 207, $n(3, k)$ should be $n(h, 3)$ and the headings of the last three pairs of rows in Table 1 should be h and $n(h, 3)$, not k and $n(3, k)$. Härtter, in his review (Zbl. 432.10032) of this paper, notes these misprints and says that the conjecture (p. 207) of the present writer was effectively proved by Hofmeister (1968) apart from a

TABLE 1. Known Congruent (C) and Noncongruent (N) Numbers less than 1000. The entry for $a = 40c + r$ is in column c and row r.

r \ c =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24 = c / r
1	NB	C1	□	□	CG	N9	N1	N1	N9	□	N1	□	NJ	N1	CG	N1	N1	N9	C&	C1	□	□	1	N9	□	1
2	NB	NB	N2	NX	□	NX	□		NJ	NX	NJ	CG		□	N2	CG	NJ	NJ	□	NJ		NX	□	NX	NL	2
3	N3	N3	N3	N&	N3	NJ	□	N3	C&	□	NJ	N3	NJ	N3	N3	□	N3	N3	CJ	NJ	NJ	NJ	N3	NJ	□	3
4	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	4
5	C5	□	CG	□	CG	CG	□	C&	□	CJ	□	C&	CJ	□	C&	□	C&	CJ	□	□	C&	□	CJ	□		5
6	C6	C6	C6	□	C6	C6	C&	CA	C6	C&	CJ	C6	□	C6	C6	CJ	CG	□	□	C6	CG	□	C6	C6	CG	6
7	C7	C7	CG	C7	C7	□	CJ	C&	CJ	C7	CJ	CJ	C7	C&	□	C7	C7	CJ	C7	CJ	CJ	□	C7	□	C7	7
8	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	8
9	□	□	N1	N9	□	N9	N9	□	NJ	□	N1	N1	N9	□	1	C&	N9	C&	□	N1	1	N9	C&	N1	NJ	9
10	NX	□	□	NL	N&	CA	□	NL	CA	NL	CG	□	□	NL	NJ	NL	□	NJ	NJ	NJ	□	□	CG	NJ	NJ	10
11	N3	NB	NB	N3	□	N3	N3	CG	N3	CG	NJ	NJ	N3	□	N3	NJ	CG	N3	C&	NJ	N3	NJ	□	□	N3	11
12	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	12
13	C5	C5	CG	CJ	C5	CJ	CJ	C5	□	C5	CJ	CJ	CJ	C&		C5	C5	□	C5	C5	C&	C5			CJ	13
14	C6	□	C6	C6	CG	C6	C6	□	C6	CJ	□	C6	CJ	CJ	C&	C6	CJ	C6	C6	□	C&	CJ	CJ	C6	C&	14
15	CA	CG	CG	□	□	C&	CG	CJ	CJ	□	CJ	C&	□	C&	□	C&	CJ	CJ	□	□	CJ	□	CJ	C&	□	15
16	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	16
17	N1	N9	N1	C1	N9	NJ	C1	□	N1	NJ	N9	C1	NJ	N9	1	N1	□	NJ	N9	C&	N9	1		N1	N1	17
18	□	NX	□	CG	N2	NX	NJ	NX	□	□	NJ	NX	NJ	NX	□	NJ	C&	NX	□	NX	N2	NJ			NJ	18
19	N3	N3	□	N3	N3	C&	NJ	CG	NJ	N3	N3	□	N3	□	NJ	N3	N3	NJ	N3	NJ	□	N3	NJ		NJ	19
20	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	20
21	CA	C5	C5	C&	C5	CA	□	CJ	CJ	CJ	C5	C5	CJ	C5	CJ	□	C5	C5	CG	CJ	C5	CJ	C&	C5	□	21
22	C6	C6	CG	C6	C&	CJ	C6	C6	□	C6	C6	CG	C6	C6	C&	C6	C6	□	CJ	CJ	CJ	C6	CJ	CJ	C6	22
23	C7	□	C7	C&	CJ	C7	C7	CJ	□	C7	□	C7	C7		CG		C&	CJ	C7	□	C7	C7	C&	C&	C7	23
24	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	24
25	□	CA	NJ	CG	NJ	□	CG	NJ	NJ	NJ	□	CG	C&	NJ	□	□	NJ	NJ	NJ	NJ	□	NJ	C&	□	C&	25
26	NX	NB	NX	N2	N&	C&	NJ	□	NX	C&	C&	N2	NJ	CA	NX	N2	□		NX	NJ	NJ	C&	NJ	NJ	NJ	26
27	□	N3	N3	□	N&	N3	NJ	N3	N3	□	NJ	N3	□	N3	N3		NJ	NJ	□	N3	N3	□	N3	N3	C&	27
28	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	28
29	C5	CG	C5	C5	□	C5	C5	CJ	C5	C5	CA	CJ	C5	□	CJ	C&	C&	C5	CJ	CJ	C5	CJ	□	C&	CJ	29
30	CA	CA	CA	□	CA	CJ	□	CG	□	CA	CJ	CG	CG	□	CJ	□	C&	C&	□	CJ	CJ	C&	C&	□	□	30
31	C7	C7	CG	C7	C7	CA	C7	C7	□	C&	C7	CJ	C&	CJ	CJ	C7	C&	□	C7	C&	CJ	CJ	C7	C&	C7	31
32	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	32
33	N9	N1	N1	□	N1	N1	NJ	C1	C1	N9	N1	N9	□	NJ	1	N9	N1	NJ	N9	C&	□	□	N9	1	N9	33
34	CA	NX	N&	CA	C&	□	N2	NX	NJ	NX	CG	NJ	C&	NX	□	NX	C&	NJ	NL	NX	NJ	NJ	N2	□	NJ	34
35	NB	□	N&	NJ	NJ	NJ	□	□	NJ	C&	NJ	□	NJ	NJ	NJ	NJ	□	NJ	NJ	NJ	NJ	□	C&	NJ	C&	35
36	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	36
37	C5	CG	□	C5	C5	CJ	C5	C5	CG	C5	CJ	□	CG	C5		□	C5		C5	C5	□	C5		C&	C5	37
38	C6	CG	C6	C6	□	CJ	C6	C&	C6	C6	CG	C6	C&	□	CJ	CJ	CJ	C6	C6	CG	C6	C6	□	C6	C6	38
39	CG	C7	C&	C&	C7	C7	□	C&	C7	C&	C7	C7	CJ	CJ	C7	□	CJ	C7	CG	C&	C7	C&	C7	C&	□	39
40	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	40
r \ c =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24 = c / r

finite amount of numerical calculation. This is now covered in the case of $n(h, 3)$ by the work of Bate, quoted in the paper, and by that of Rödseth (1978, 1979), Selmer (1980), and Wagstaff (1979). Part of Selmer's autorreferat is "For $k = 3$, the global problem was solved by Hofmeister in 1968, and an algorithm for the local problem has recently been given by Rödseth, utilizing a connexion (discovered by Meures) between the postage stamp problem and 'coin exchange problem' of Frobenius."

Graham and Sloane (1980) have written an important paper surveying this and related areas, in which they compare and contrast four pairs of covering and packing versions of extremal basis and "B_2-sequence" problems, half of them in modular form. They also introduce harmonious graphs, a welcome change from graceful ones. A connected graph with v vertices and $e(\geq v)$ edges is **harmonious** if the vertices can be labelled with distinct integers so that the sums of the labels at the ends of each edge form a complete system of residues (mod e). Trees, for which $e = v - 1$, are included by allowing just one vertex label to be duplicated. It is conjectured that all trees are harmonious as well as graceful.

Andrzej Mąkowski writes to say that the conjecture made by Erdős at the end of his note [1980, 391], that $\Sigma(1/p_i) = \infty$ implies $b_{i+1} - b_i = 1$ infinitely often, is not true, since we may take $\{p_i\}$ to be the set of primes $\equiv 1 (\mathrm{mod}\ a)$, giving $b_{i+1} - b_i \geqslant a$ for all i, since the b_i (whose prime factors all belong to the chosen set) are all $\equiv 1 (\mathrm{mod}\ a)$.

In asking for the number, $L(i, j, m, n)$, of i-j reduced latin rectangles, Hamilton and Mullen [1980, 392] speculated that $(n-m)!L(1,0,m,n) = (n-1)!L(1,1,m,n)$. James Nechratal says that this becomes immediately clear on rewriting it as $L(1,0,m,n) = (n-1)_{m-1}L(1,1,m,n)$, since the number of $m \times n$ latin rectangles with first row $1,2,\ldots,n$ and the specified first column is independent of the structure of that column, and there are $(n-1)_{m-1}$ possible first columns. He also points out that credit should have been given to F. W. Light (1973) who computed $L(1,0,4,n)$ for $n \leqslant 8$.

I apologize that the article of S. Doran [1980, 474] turned out to be little more than an exercise, with a strong hint for its solution. A modification of his example affirmatively answers the question: Does there exist a Banach *-algebra without identity with no nonzero positive functionals? This was observed, in rapid succession, by Laurence J. Dixon, Joseph Szucs, R. Schlafly, H. Garth Dales, David L. Johnson, and Robert Whitley; the first three of these came a week before my copy of the MONTHLY!

John P. Mayberry and Martin Fürer, as well as Baxter himself, write to say that Baxter's note [1981, 50] is not quite correct. The facts are these (1) n digits of $\sqrt{2}$, and indeed of any algebraic number, can be computed in time $O(n \ln n \ln\ln n)$; (2) n digits of π, and of any number based on the arithmetic-geometric sequence, can be computed in time $O(n(\ln n)^2 \ln\ln n)$; (3) e can also be computed in this time. For details of the two latter statements, see Brent (1976). Baxter observes, however, that the unsolved problem remains: is there an efficient computation of *e expressed as an iterative sequence*? Baxter's note was concerned with sequences of the form $x_t = F(x_{t-1}, x_{t-2}, \ldots)$ where F uses arithmetic operations, so that his claim that no equally efficient method of computing e is known, was intended to mean "no equally efficient method expressed as an iterative sequence." Brent's method uses *iteration* but does not appear to be equivalent to an iterative sequence since his computation of $U(x)$ is defined by $g(x, \pi, f(x))$ where f and g are expressible as iterative sequences, but the computation of e then involves the solution of $U(x) = 1$.

I thank the many correspondents mentioned in this article and many more who are not. Only with their help can this useful pooling of information take place.

References

M. Ajtai, V. Chvátal, M. M. Newborn, and E. Szemerédi, Crossing-free Hamiltonian circuits, Ann. Discrete Math. (to appear).

Dale E. Alspach, A fixed point free nonexpansive map, Proc. Amer. Math. Soc. (to appear).

J.-C. Bermond, Graceful graphs, radio antennae and French windmills, in Graph Theory and Combinatorics, Pitman, London, 1979, pp. 18–37.

Norman Biggs, Some odd graph theory, Proc. 2nd Internat. Conf. Combin. Math., New York, 1978; Ann. N.Y. Acad. Sci., 319 (1979) 71–81; MR 81c: 05026.

Richard P. Brent, Fast multiple-precision evaluation of elementary functions, J. Assoc. Comput. Machinery, 23 (1976) 242–251.

F. M. Dekking, Strongly nonrepetitive sequences and progression-free sets, J. Combin. Theory Ser. A, 27 (1979) 181–185.

P. Erdős and R. L. Graham, Old and New Problems and Results in Combinatorial Number Theory, Monographies de l'Enseignment Math. no. 28, Geneva, 1980.

Jean Françon and Gérard Viennot, Permutations selon leurs pics, creux, double montées et double descentes, nombres d'Euler et nombres de Genocchi, Discrete Math., 28 (1979) 21–35.

Cyril W. L. Garner and Agnes M. Herzberg, On McCarty's queen squares, this MONTHLY, 88 (1981), 612–613.

Basil Gordon, Robert M. Guralnick, and Michael D. Miller, On cyclic commutator subgroups, Aequationes Math., 17 (1978) 112–113.

R. L. Graham and N. J. A. Sloane, On additive bases and harmonious graphs, SIAM J. Alg. Discrete Math., 1 (1980) 382–404.

Richard K. Guy, Unsolved Problems in Intuitive Mathematics, vol. I, Number Theory, Springer-Verlag, New York, 1981.

Joseph Hammer, Unsolved Problems Concerning Lattice Points, Research Notes in Math. 16, Pitman, London, 1977; Zbl. 355.52001.

______, Lattice points and area-diameter relations, Math. Mag., 52 (1979) 25–26; MR 80d: 10044.

Agnes M. Herzberg and Cyril W. L. Garner, Latin queen squares, Utilitas Math. (to appear).

Gerd Rolf Hofmeister, Asymptotische Abschätzungen für dreielementige Extremalbasen in natürlichen Zahlen, J. Reine Angew. Math., 232 (1968) 77–101; MR 38 #1068; Zbl. 165, 62.

Charlotte Huang, Anton Kotzig, and Alexander Rosa, Further results on tree labellings, Utilitas Math. (to appear).

A. Kotzig, On certain vertex-valuations of finite graphs, Utilitas Math., 4 (1973) 261–290.

Jean Lagrange, Thèse d'Etat de l'Université de Reims, 1976.

______, Construction d'une table de nombres congruents, Bull. Soc. Math. France Mém. no. 49–50 (1977) 125–130; MR 58 #5498.

F. W. Light, A procedure for the enumeration of $4 \times n$ latin rectangles, Fibonacci Quart., 11 (1973) 241–246; MR 48 #3764; Zbl. 271.05007.

Roger C. Lyndon, Equations in groups, Bol. Soc. Brasil. Math., 11 (1980) no. 1.

Monroe Newborn and W. O. J. Moser, Optimal crossing-free hamilton circuit drawings of K_n, J. Combin. Theory Ser. B, 29 (1980) 13–26.

Š. Porubský, On MacMahon's segmented numbers and related sequences, Nieuw Arch. Wisk., (3) 25 (1977) 403–408; MR 58 #5575.

Helmut Prodinger and Friedrich J. Urbanek, Infinite 0-1-sequences without long adjacent identical blocks, Discrete Math., 28 (1979) 277–289.

Raphael M. Robinson, Multiple tilings of n-dimensional space by unit cubes, Math. Z., 166 (1979) 225–264.

Öystein J. Rödseth, On a linear diophantine problem of Frobenius, I, II, J. Reine Angew. Math., 301 (1978) 171–178; 307/308 (1979) 431–440; Zbl. 374.10011, 395.10021.

D. G. Rogers, A multiplication theorem for regular perfect systems of difference sets (preprint).

Ernst S. Selmer, On the postage stamp problem with three stamp denominations, Math. Scand., 47 (1980) 29–71; Zbl. 436.10027.

Volker Strehl, Alternating permutations and modified Ghandi-polynomials, Discrete Math., 28 (1979) 89–100.

Jean Turgeon, An upper bound for the length of additive sequences of permutations, Utilitas Math., 17 (1980) 189–196.

Samuel S. Wagstaff, Additive h-bases for n, Proc. Number Theory Conf., Carbondale, 1979, Springer Lecture Notes in Math. 751 (1979) 302–327; Zbl. 418.10051.

CLASSROOM NOTES

EDITED BY J. ARTHUR SEEBACH, JR.

Material for this department should be sent to Professor J. Arthur Seebach, Jr., Department of Mathematics, St. Olaf College, Northfield, MN 55057.

MORE ON SIMILARITY OF MATRICES

RICHARD A. DAVIS

Department of Mathematics and Computer Science, Manhattan College, Bronx, NY 10471

William Watkins, in [**1**], has given an elementary proof of the following:

THEOREM. *Let A and B be real n-by-n matrices. If A is similar to B over the complexes, then A is similar to B over the reals.*

Those familiar with the polynomial invariants of an endomorphism should have little difficulty in proving this result for arbitrary fields, although the method of proof is quite different from that in [**1**]. See, for example, [**2**, pp. 305–309]. The purpose of this note is to indicate how Watkins's

technique can be modified to handle more general situations, such as those given in the theorem below, where the theory of polynomial invariants does not apply.

The following facts are assumed: First, if the determinant of a matrix is nonzero, then the matrix is invertible. Second, if $f(x_1, x_2, \ldots, x_m)$ is a nonzero polynomial over an infinite field K, then there exist $k_1, k_2, \ldots, k_m$ belonging to K such that $f(k_1, k_2, \ldots, k_m) \neq 0$. See [**2**, pp. 120–123]. Finally, there exists a basis for any vector space over a field, finitely generated or not. See [**2**, pp. 85–87].

DEFINITION. Suppose $\{A_j\}_{j\in J}$ and $\{B_j\}_{j\in J}$ are sets of n-by-n matrices over a field K. We say that these sets are similar over K if there is an invertible n-by-n matrix S over K such that $A_j = S^{-1}B_jS$ for all j in J.

THEOREM. *Suppose K is an infinite field and L is any field containing K. Let $\{A_j\}_{j\in J}$ and $\{B_j\}_{j\in J}$ be sets of n-by-n matrices over K. If these sets are similar over L, then they are similar over K.*

Proof. Suppose that $A_j = S^{-1}B_jS$ for all j in J, where S is an invertible matrix over L. We can write $S = P_1h_1 + P_2h_2 + \cdots + P_mh_m$, where the h_i are elements of L, linearly independent over K, and the P_i are n-by-n matrices over K for $i = 1, 2, \ldots, m$. Then $SA_j - B_jS = 0$ implies that the sum $(P_1A_j - B_jP_1)h_1 + (P_2A_j - B_jP_2)h_2 + \cdots + (P_mA_j - B_jP_m)h_m = 0$, and so $P_iA_j = B_jP_i$ for each i and j.

$$f(x_1, x_2, \ldots, x_m) = \det(P_1x_1 + P_2x_2 + \cdots + P_mx_m)$$

is a polynomial over K and is not identically zero since $f(h_1, h_2, \ldots, h_m) = \det S \neq 0$. Thus, there are elements $k_1, k_2, \ldots, k_m$ in K such that $f(k_1, k_2, \ldots, k_m) \neq 0$, and the matrix $M = P_1k_1 + P_2k_2 + \cdots + P_mk_m$ is invertible over K. Since $MA_j = B_jM$ for all j in J, we are finished.

This proof carries over unchanged if K is finite with more than n elements, and with slight modifications we can prove the theorem when K has n or fewer elements. See [**3**, pp. 198–202], whose treatment we have followed here.

The notion of similarity of sets of matrices and the theorem proved above can be applied, for example, in situations involving the simultaneous triangulation or diagonalization of a family of commuting matrices. See [**4**, p. 206].

References

1. W. Watkins, Similarity of matrices, this MONTHLY, 87 (1980) 300.
2. S. Lang, Algebra, Addison-Wesley, Reading, Mass., 1965.
3. C. W. Curtis and I. Reiner, Representation Theory of Finite Groups and Associative Algebras, Interscience, New York, 1962.
4. K. Hoffman and R. Kunze, Linear Algebra, Prentice-Hall, Englewood Cliffs, N.J., 1961.

PROBLEMS AND SOLUTIONS

*Send all **proposed** problems, in duplicate if possible, to Professor Vladimir Drobot, Department of Mathematics, University of Santa Clara, Santa Clara, CA 95053. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the addresses given at the head of each problem set.*

A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

ELEMENTARY PROBLEMS

Solutions of these Elementary Problems should be mailed in duplicate to Dr. J. L. Brenner, 10 Phillips Road, Palo Alto, CA 94303 (USA), by April 30, 1982. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label (for acknowledgment).

E 2914. *Proposed by R. C. Lyness, Southwold, England.*

A circle B lies wholly in the interior of a circle A. S is the set of all circles each of which touches B externally and A internally.

(i) Find the locus of the internal center of similitude of the pairs of circles from S. (ii) Prove that every point of the locus, except one, is the i.c.s. of exactly one pair of circles from S.

E 2915. *Proposed by Leo J. Alex, SUNY at Oneonta, N.Y.*

Let k be fixed. It is conjectured that $3^n = \Sigma_{i=1}^{k} 2^{r_i}$ has only finitely many integral solutions $(n, r_1, r_2, \ldots, r_k)$. This is clear if $k = 1$; Størmer proved it if $k = 2$, and S. S. Pillai if $k = 3$. Prove the assertion for $k = 4$.

E 2916. *Proposed by R. Sivaramakrishnan, University of Calcutta, India.*

Let $\phi_2(r)$ represent the number of integers $a, 1 \leqslant a \leqslant r$, with $(a, r) = (a+1, r) = 1$. Prove that, for $r, n \geqslant 1$, the relation

$$\Sigma\phi(nr/d^2)\, d\mu(d) = \phi(n/u)\phi(r/u)\phi_2(u)$$

holds. Summation extends over all $d, d\,|\,(n, r)$. u is the greatest common square-free unitary divisor of n and r. (A divisor d of c is called unitary if $(d, c/d) = 1$.) ϕ [μ] is the Euler totient [Möbius function].

E 2917. *Proposed by F. W. Luttman, California State University, Sonoma.*

Let P_0 be a convex polygon of n sides and let $0 < f < 1$. Let $P_0, P_1, P_2, \ldots$ be a nested sequence of polygons similar to P_0 with the following properties:

(1) P_{k+1} is a linear contraction of P_k by the factor f.

(2) Two adjacent sides of P_{k+1} lie on P_k. (Necessarily P_k and P_{k+1} share a single vertex.)

(3) The vertex which P_k shares with P_{k+1} lies next clockwise from the vertex it shares with P_{k-1}.

There is precisely one point lying inside all P_k's. Construct it. (See H. S. M. Coxeter, *Introduction to Geometry*, p. 164.)

E 2918. *Proposed by Jordi Dou, Barcelona, Spain.*

Show that an isosceles triangle can be dissected symmetrically around the principal median into seven acute isosceles triangles except when the vertex angle A is $90°, 120°$, or when $135° \leqslant A \leqslant 144°$.

E 2919. *Proposed by Elgin Johnston, Iowa State University, Ames.*

(a) Let $\{a_k\}_{k=1}^{\infty}$ be a sequence of real numbers. Suppose that, for each prime p, the sequence $\{a_{np}\}_{n=1}^{\infty}$ is increasing and bounded above. Let $\{b_k\}$ be the sequence obtained when those elements with prime subscript are deleted from $\{a_k\}$. (So $b_1 = a_1$, $b_2 = a_4$, $b_3 = a_6$, $b_4 = a_8$, $b_5 = a_9, \ldots$.) Show $\lim_{k\to\infty} b_k$ exists.

(b) Suppose in part (a) the phrase "is increasing and bounded above" is replaced by "converges." Does the result still hold?

SOLUTIONS OF ELEMENTARY PROBLEMS

Sum of Number of Divisors $\Sigma d(k)$, $k \leqslant n$

E 2780 [1979, 503]. *Proposed by Jim Totten, University of Saskatchewan.*

Let $d(n)$ be the number of (positive integral) divisors of the natural number n and define $S(n)$ as $\Sigma d(k)$, with the sum taken over all divisors k of n. Determine the values of n for which $n = S(n)$.

Solution by Arnold Adelberg, Grinnell College, and Lee Erlebach, Michigan Technological University, independently. The solutions are $n = 1, 3, 18, 36$. First suppose $n = p^k$ is a prime power > 1. Then $(*)$ $S(p^k) = (k+1)(k+2)/2$, so that $n = 3$ if $n = S(n)$. Now S is a multiplicative function since d is multiplicative. Therefore, if $n = S(n)$ and $n \neq 1, 3$, there are prime powers $P, Q, P \mid n, Q \mid n$, such that $(P, n/P) = (Q, n/Q) = 1$, $S(P)/P > 1$, and $S(Q)/Q < 1$. From $(*)$, we see by an easy computation that P must be 2, 4, or 8. Thus $1 < S(P)/P \leqslant 3/2$; so Q is not a power of 2 and satisfies $2/3 \leqslant S(Q)/Q < 1$. Again from $(*)$, the only possible Q is 9. Hence $S(Q)/Q = 2/3$, $S(P)/P = 3/2$, and no prime other than 2 or 3 may occur in the prime power factorization of n, establishing our result.

Also solved by H. L. Abbott (Canada), Trygve Breiteig (Norway), Bela Brindza (Hungary), Robert Breusch, Duane M. Broline, Ken Brown, Thomas E. Elsner, Joe Flowers, Lorraine L. Foster, Ralph P. Grimaldi, Carl Hurd, Thomas Jager, Lenny Jones, Michael Josephy (Costa Rica), Bela Kis (Hungary), Jonathan Leech, Bernard L. Martin, Joel Matkin, Victor Pambuccian (Romania), Richard Quindley & Philip Scalisi & Thomas Moore, Jeffrey Shallit, James Theiler, John T. Ward, Kenneth L. Yocom, and the proposer.

The Determinant of the Generating Function for Distances in a Tree

E 2827 [1980, 303]. *Proposed by Gerard Letac, Université Paul Sabatier, Toulouse, France.*

The n vertices of a tree are labeled with the integers $1, 2, \ldots, n$. Let $d(i, j)$ denote the number of edges between i and j. Compute the determinant of the $n \times n$ matrix with (i, j) element $x^{d(i,j)}$.

Solution by J. D. Primer, student, Columbia High School, Maplewood, N.J., and D. Wolland, student, Hunter College High School, N.Y. Note $\Delta_1 = 1$, $\Delta_2 = (1 - x^2)$. We prove that $\Delta_n = (1 - x^2)^{n-1}$. In a tree T on $n > 2$ vertices, assume without loss of generality that vertex 1 has degree 1 and that $d(1, a) = 1$. To compute Δ_n, subtract x times row a from row 1. The new values in row 1 are $(1 - x^2)\delta_{1j}$ $(j = 1, 2, \ldots, n)$. Thus $\Delta_n = (1 - x^2)\Delta_{n-1}$.

Also solved by Steven Andrianoff, Kenneth Bernstein, Irl Bivens, N. J. Fine, Ira Gessel, O. P. Lossers (Netherlands), Joel Levy, Tim McMillan, N. Miku (Netherlands), R. K. Oliver, H. Prodinger, Santa Clara Problem Solving Integral Domain, Robert Singleton, B. Viswanathan (Canada), Michael Woltermann, and the proposer.

Submatrices of Real Symmetric Orthogonal Matrices

E 2840 [1980, 489]. *Proposed by Felix T. Smith, SRI International.*

A real symmetric orthogonal matrix is to be constructed when one of the rows, say $C_{1j} = \alpha_j$,

with $\Sigma\alpha_j^2 = 1$, is specified. For $n > 2$, there are 2 and only 2 solutions; give a general expression for the (real) C_{ij} in terms of the α_j. What happens when the matrix is symmetric and unitary—when the C_{ij} are complex and $C_{1j} = \alpha_j + i\beta_j$?

6303 [1980, 495]. *Proposed by the editors*. (See p. 707.)

Under what conditions on A can the leading principal $t \times t$ minor of the real orthogonal matrix A be replaced by a symmetric $t \times t$ minor so that the property of orthogonality is preserved?

Solution to both problems by E. G. Straus, UCLA. The statement of Problem E2840 contains its own refutation. For, if the choice of the first row could be an arbitrary unit vector and then determine the rest of the matrix A in a 2-valued way, then choosing $\alpha_1 = 1, \alpha_2 = \cdots = \alpha_n = 0$, we only have to choose an $(n-1) \times (n-1)$ symmetric orthogonal matrix to complete the remaining rows and columns, and according to the statement this can be done in an $(n-2)$-parametric way.

A simple counting of dimensions shows that there is an $[n^2/4]$-dimensional family of $n \times n$ symmetric orthogonal matrices. Thus $n-1$ parameters obviously do not suffice to determine its elements up to a finite number of choices. To see this, observe that a symmetric matrix is unitarily equivalent to a real diagonal matrix and, since our matrix is orthogonal, it is equivalent to a diagonal matrix with entries ± 1. Conversely, any orthogonal matrix with all eigenvalues real is symmetric. Thus the matrix determines two orthogonal subspaces $\mathbb{R}^k$ and $\mathbb{R}^{n-k}$ of $\mathbb{R}^n$ so that it acts as the identity on $\mathbb{R}^k$ and minus the identity on $\mathbb{R}^{n-k}$. The centralizer in $O(\mathbb{R}^n)$ of such a diagonal matrix is $O(\mathbb{R}^k) \times O(\mathbb{R}^{n-k})$ so that the set of distinct conjugates has dimension

$$\begin{aligned}\dim O(\mathbb{R}^n) - \dim O(\mathbb{R}^k) - \dim O(\mathbb{R}^{n-k}) &= \binom{n}{2} - \binom{k}{2} - \binom{n-k}{2} \\ &= [n^2/4] \quad \text{if} \quad k = [n/2].\end{aligned}$$

This analysis also enables us to give the correct answer to Problem E2840. Let r_1 be the given first-row vector and let the column vector $e_1 = (1, 0, \ldots, 0) = e_1^+ + e_1^-$, $e_1^+ \in \mathbb{R}^k$, $e_1^- \in \mathbb{R}^{n-k}$. Then

$$r_1 = Ae_1 = e_1^+ - e_1^- .$$

Thus the choice of r_1 determines $e_1^+ = (e_1 + r_1)/2$, $e_1^- = (e_1 - r_1)/2$. We are now free to complete the choice of $\mathbb{R}^k$ and $\mathbb{R}^{n-k}$ leading to a symmetric orthogonal matrix A with prescribed first row.

In Problem 6303, since the rows $r_{t+1}, \ldots, r_n$ are to remain the same, it is clear that the only transformation is an orthogonal transformation T on the space $\mathbb{R}^t$ spanned by the rows $r_1, \ldots, r_t$. This T must leave the columns $c_{t+1}, \ldots, c_n$ of the $n \times t$ matrix $[r_1^*, \ldots, r_t^*]^*$ fixed. In other words T is a unitary transformation of $\mathbb{R}^t$ which acts as the identity on the space C spanned by $c_{t+1}, \ldots, c_n$. The $t \times t$ submatrix A_t of A can therefore be replaced by a symmetric one if and only if

$$A_t = TS$$

where S is symmetric and T is $t \times t$ orthogonal acting as the identity on C.

For example, if $n = 3$, $t = 2$, the only orthogonal transformation leaving the column $[a_{31}, a_{32}]^*$ fixed is the identity, unless $a_{31} = a_{32} = 0$. Thus either A_2 is already symmetric, or $a_{31} = a_{32} = a_{13} = a_{23} = 0$, $a_{33} = \pm 1$ and A_2 can be replaced by an arbitrary symmetric 2×2 matrix $[\cos\theta, \sin\theta; \sin\theta, -\cos\theta]$.

E 2340 also solved by F. S. Cater, A. C. Hindmarsh, H. Kestelman (United Kingdom), A. Nijenhuis, R. L. Pegs, and M. J. Sherman.

A Condition that Implies lim $f(n)/n = \inf f(n)/n$

E 2841. *Proposed by J. Michael Steele, Stanford University.*

Let $f(n)$ be a real valued function defined for every natural number n. Suppose $f(a+b+c) \leqslant f(a)+f(b)+f(c)$ for all a, b, c such that a/b and b/c are between $1/3$ and 3. Show that $\lim_{n\to\infty} f(n)/n$ exists and equals $\inf_n f(n)/n$.

Solution by J. Muldowney, Edmonton, Canada. This result will be proved subject to the weaker hypothesis that, for some $\varepsilon > 0$, $f(2a+b) \leqslant 2f(a)+f(b)$ if $a \leqslant b \leqslant a(1+\varepsilon)$.

Let $g(x) = f(x)/x$. Then $z = 2x+y$, $x \leqslant y \leqslant x(1+\varepsilon)$, imply

$$g(z) \leqslant [2xg(x) + yg(y)]/z. \tag{1}$$

Choosing $x = y = 3^{n-1}w$, (1) implies that $g(3^n w)$ is nonincreasing for each w and, further, if $3/w < \varepsilon$, (1) shows that $\max\{g(z): 3^n w \leqslant z \leqslant 3^{n+1}w\}$ is nonincreasing in n. Therefore $-\infty \leqslant \sigma < \infty$, where $\sigma = \limsup_{n\to\infty} g(n)$. Since the case $\sigma = -\infty$ is trivial, we may assume $\sigma = 0$, because $f(x) - \sigma x$ satisfies the same hypothesis as $f(x)$.

To complete the proof, it remains to show that $\inf_n\{g(n)\} = 0$. Suppose $g(w) < 0$ for some w. We assert that, for each $k \geqslant 0$,

$$0 > \limsup_{z\to\infty} g(z), \qquad z \in \bigcup_{n=0}^{\infty} \left[3^n w, 3^n(w+k)\right]. \tag{2}$$

But the case $k = 2w$ implies $\sigma < 0$ contradicting $\sigma = 0$.

The assertion $(2)_k$ is proved by induction on k. Since $g(3^n w)$ is nonincreasing, $(2)_0$ holds. If $(2)_k$ holds then, in particular, $g(3^N(w+k)) < 0$ for all large N. If $z \in [3^n(w+k), 3^n(w+k+1)]$, then $z = 2x+y$ where $x = 3^{n-1}(w+k)$, $y \in [3^{n-1}(w+k), 3^{n-1}(w+k+3)]$ and (1) together with $\sigma = 0$ implies

$$0 > \frac{2(w+k)}{(w+k+3)} g\left(3^N(w+k)\right) \geqslant \limsup_{z\to\infty} g(z), \qquad z \in \bigcup_{n=0}^{\infty} \left[3^n(w+k), 3^n(w+k+1)\right].$$

This inequality with $(2)_k$ implies $(2)_{k+1}$.

The same sort of argument proves the assertion: If $f(\Sigma a_i) \leqslant \Sigma f(a_i)$ for all natural numbers a_i (spaced not too far apart) then $\lim f(n)/n = \inf f(n)/n$.

Also solved by A. Meir (Canada), A. Smuckler (Israel), and the proposer.

Symmetric Difference of Areas of a Triangle and a Circle

E 2842 [1980, 577]. *Proposed by Jordi Dou, Barcelona, Spain.*

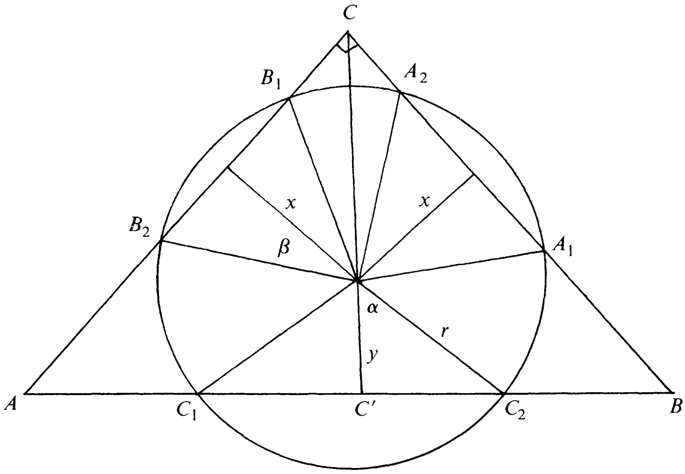
Let T be an isosceles right triangle. Let S be the circle such that the difference between the areas $T \cup S$ and $T \cap S$ is minimal. Show that the center of S divides the altitude on the hypotenuse of T in the golden ratio.

Solution by the Proposer. First we show the following—which hold incidentally for any triangle T:

1. Half the circumference of S lies within T and half without.
2. The chords of S lying along the sides of T are in proportion to the sides of T.

In the following, when X is a point set in the plane, $[X]$ denotes its area. Let $\overrightarrow{A_1A_2}$ be the counterclockwise chord of A lying opposite vertex A, as in the figure. $\overrightarrow{B_1B_2}$ and $\overrightarrow{C_1C_2}$ are similarly defined.

Proof of 1. Let S have fixed center. As its radius increases differentially, $[T - S]$ decreases in proportion to the total length of the arcs of S lying within T, and $[S - T]$ increases in proportion to the total length of the arcs of S lying without T. Thus for $\Delta = [T \cup S] - [T \cap S] = [T - S] + [S - T]$ to be instantaneously unchanging, these arcs must be equal.



Proof of 2. Let S have fixed radius. Since $[T - S] - [S - T] = [T] - [S]$, and the latter is constant, it suffices to minimize $[S - T]$ (or $[T - S]$) in order to minimize Δ.

Let the center of S be displaced by $\mathbf{u}dx$ where $\mathbf{u}$ is a unit vector. The change in $[S - T]$ along $\overrightarrow{A_1A_2}$ can be represented by $\mathbf{u}dx \times \overrightarrow{A_1A_z}$—this vector is normal to the plane of T and its direction determines the sign of the change. Similarly for the other two sides. Hence the total change in $[S - T]$ is $\mathbf{u}dx \times (\overrightarrow{A_1A_2} + \overrightarrow{B_1B_2} + \overrightarrow{C_1C_2})$. Since this quantity must be $\mathbf{0}$ for all $\mathbf{u}$ when S is the minimizing circle, we require $\overrightarrow{A_1A_2} + \overrightarrow{B_1B_2} + \overrightarrow{C_1C_2} = 0$. This condition is equivalent to the assertion of 2.

Let us normalize T by making $BC = CA = 1$, and $AB = \sqrt{2}$. Let y and x be as in the figure. Note that

$$y = \frac{1}{2}\sqrt{2} - \sqrt{2}x \quad \text{or} \quad 2y^2 = (1 - 2x)^2. \tag{1}$$

From Assertion 1, $\alpha + 2\beta = \pi/2$, so that $\cos\alpha = \sin 2\beta$; thus

$$y/r = \frac{2x}{r^2}\sqrt{r^2 - x^2} \quad \text{or} \quad y^2r^2 = 4x^2r^2 - 4x^4. \tag{2}$$

From Assertion 2,

$$\sqrt{r^2 - y^2}\Big/\sqrt{2} = \sqrt{r^2 - x^2}\Big/1,$$

so $2x^2 - y^2 = r^2$ and eliminating y using (1), we have

$$2r^2 = 4x - 1. \tag{3}$$

Eliminating y and r from (2) by means of (1) and (3), we have $16x^4 - 16x^3 - 12x^2 + 8x - 1 = 0$, which factors into $(4x^2 + 2x - 1)(4x^2 - 6x + 1) = 0$. (This is more easily seen by replacing $2x$ by X.) Of the four solutions the one which applies is clearly $x = \frac{1}{4}(\sqrt{5} - \frac{1}{2})$. (One is negative, another exceeds 1, and the remaining one makes r negative.) Thus $r = \sqrt{\frac{1}{2}(\sqrt{5} - 2)}$ and $y = \frac{1}{4}\sqrt{2}(3 - \sqrt{5})$. We have finally

$$\frac{OC}{OC'} = \frac{\sqrt{2}\,x}{y}\,\frac{\sqrt{5} + 1}{2},$$

which is the golden ratio.

As an alternative, the last portion of the proof can be done as follows: Let $\tau = \sqrt{2}x/y =$

OC/OC'. From (1)

$$2x = \tau(1-2x) \quad \text{or} \quad x = \frac{1}{2}\frac{\tau}{\tau+1}. \tag{4}$$

From (2)

$$2r^2 = 4r^2\tau^2 - 4x^2\tau^2;$$

and using (3)

$$4x - 1 = 2\tau^2(4x-1) - 4x^2\tau^2;$$

and then (4)

$$\frac{\tau-1}{\tau+1} = 2\pi^2\frac{\tau-1}{\tau+1} - \frac{\tau^4}{(\tau+1)^2}.$$

Multiplying by $(\tau+1)^2$ we obtain $\tau^4 - 3\tau^2 + 1 = 0$. Hence τ, being positive, is the Golden Ratio, or its reciprocal. But in the latter case $r^2 < 0$.

Also solved by St. Olaf's College Problem Solving Group.

Iterates of an Affine Transformation

E 2846 [1980, 577]. *Proposed by D. Wiedemann, Institute for Defense Analyses.*

Let V be an n-dimensional vector space ($n > 0$) over a field with characteristic $p \neq 0$. Let A be any affine map (linear plus a constant) from V to itself. Show there is an $x \in V$ and a positive integer $k \leqslant np$ such that the kth iterate of A takes x to itself.

Solution by N. Miku, Catholic University, Nijmegen, The Netherlands. Put $Ax = Bx + a$, B linear. Then $A^kx = B^kx + C_ka$, where $C_k = B^{k-1} + \cdots + B + I$. Note that $(B-I)C_k = B^k - I$. Now $x = A^kx$ is equivalent to $(I - B^k)x = C_ka$, which in turn is equivalent to $(B-I)x + a \in \operatorname{Ker} C_k$. A solution exists iff $a \in \operatorname{Ker} C_k + \operatorname{Im}(B-I)$. We suppose that $B - I$ is singular, since otherwise we could take $k = 1$. Now the restriction of B to the invariant space S belonging to the eigenvalue 1 can be written as $I + N$, N nilpotent. Thus

$$C_{k|S} = \sum_{i=0}^{k-1}(I+N)^i = \sum_{i=0}^{k-1}\sum_{j=0}^{i}\binom{i}{j}N^j = \sum_{j=0}^{k-1}\left(\sum_{i=j}^{k-1}\binom{i}{j}\right)N^j = \sum_{j=0}^{k-1}\binom{k}{j+1}N^j.$$

If k is a multiple of p, this equals N^{k-1}, and if $k - 1 \geqslant \dim S$ this is 0. Thus the invariant space S is in the kernel of C_k if $k = lp \geqslant 1 + \dim S$. Moreover $\operatorname{Im}(B-I) \supset \operatorname{Im}(B-I)^n$, which is a complement of S. Thus we can take $k = (n-1)p$ if $(n-1)p \geqslant \dim S + 1$. If not, then $n = 1$, or $n = 2$, $p = 2$ and $\dim S = 2$. But if $n = 1$ clearly $k = p$ is a solution, and if $n = 2$, $p = 2$ we can take $k = 2$ if $\dim S = 1$ or $B = I$, and $k = 4$ otherwise. Thus if $k = np$ there is always a solution. In fact the only cases where this is the smallest possible value of k are $n = 1$, and $n = 2$, $p = 2$, $B \neq I$ and $\dim S = 2$; in all other cases $k = (n-1)p$ gives a solution.

Also solved by E. Badertscher and the proposer.

ADVANCED PROBLEMS

Solutions of these Advanced Problems should be mailed in duplicate to Professor David Borwein, Department of Mathematics, University of Western Ontario, London, Ontario, Canada N6A5B9, by April 30, 1982. The solver's full post-office address should be on each sheet.

6368. *Proposed by Robert E. Shafer, Berkeley, Calif.*

Consider the integral

$$I = \int_{v/2}^{(v+1)/2}\psi(z)\,dz = \log\left[\Gamma\left(\frac{v+1}{2}\right)\Big/\Gamma(v/2)\right], \qquad \operatorname{Re} v > 0$$

where ψ denotes the logarithmic derivative of the gamma function. Simpson's rule with two subintervals gives

$$I=\frac{1}{12}\psi\left(\frac{v}{2}\right)+\frac{1}{3}\psi\left(\frac{v+1/2}{2}\right)+\frac{1}{12}\psi\left(\frac{v+1}{2}\right)-E,$$

where

$$E=\frac{1}{92160}\psi^{(\mathrm{iv})}\left(\frac{v}{2}+\frac{\theta}{2}\right),\qquad 0<\theta<1.$$

Show that for real v, E is negative, by deriving the representation

$$E=-\int_0^1\left\{\frac{1}{\log\left(\frac{1+x}{1-x}\right)}-\frac{1+2\sqrt{1-x^2}}{6x}\right\}\left(\frac{1-x}{1+x}\right)^{v-1}\frac{dx}{1+x}.$$

6369*. *Proposed by Joseph O'Rourke, Johns Hopkins University.*

Let a_i and b_j be positive integers, $i, j = 1,2$ $(1 \leqslant a_i, b_j \leqslant N)$. Find the minimal positive value of $\Sigma a_i^{1/2} - \Sigma b_j^{1/2}$. Suggest or prove asymptotic results. Consider the corresponding problem with sums of three terms.

6370*. *Proposed by P. Erdős, Hungarian Academy of Sciences.*

For each set of positive integers $\{a_i\}$ such that $\Sigma a_i^{-1} \leqslant 1$, define $f_1(n) := \Sigma 1/\phi(a_i)$, the sum being extended over those $a_i \leqslant n$. Set $f(n) = \max f_1(n)$, the maximum being taken over the admissible sets $\{a_i\}$; ϕ is Euler's totient. Estimate $f(n)$ as well as you can.

6371. *Proposed by Bruce C. Berndt, University of Illinois, Urbana-Champaign.*

Show that

$$\lim_{N\to\infty}\left(\sum_{k=1}^{N}k^{-1/2}\operatorname{Log}k-2\sqrt{N}\operatorname{Log}N+4\sqrt{N}\right)=-\zeta(1/2)\{\pi/4+\gamma/2+(\log(8\pi))/2\}$$

where $\zeta(s)$ denotes the Riemann zeta-function and γ denotes Euler's constant. Note $\zeta(1/2) = (2^{1/2}+1)\Sigma_1^\infty(-1)^n n^{-1/2}$.

6372. *Proposed by David R. Brillinger, University of California, Berkeley.*

Let n be a positive integer and f in $L_p(R)$ with $p = (n+1)/n$. Then

$$\int\cdots\int|f(x_1)\cdots f(x_n)f(x_1+\cdots+x_n)|dx_1\cdots dx_n\leqslant\|f\|_p^{n+1}.$$

(This inequality yields directly the result of Problem 5314, this MONTHLY, 72 (1965) 795, discussed by H. Dym, this MONTHLY, 87 (1980) 53-54.)

6373. *Proposed by F. S. Cater, Portland State University.*

By an "order space" we mean a totally ordered nonvoid set endowed with the open interval topology. Let X be an order space, and let Y be either an order space or a locally compact Hausdorff space. Let f be a mapping of X into Y that is not everywhere continuous, and let G denote the graph of f, $G = \{(x, f(x)) : x \in X\}$.

(i) Prove that G is not both a closed and connected subset of the product space $X \times Y$. (Compare problem 6255 in this MONTHLY.)

(ii) Show, by example, that (i) need not hold for complete connected metric spaces X and $Y = f(X)$, if X is not an order space.

SOLUTIONS OF ADVANCED PROBLEMS

Submatrices of Real Symmetric Orthogonal Matrices

6303 [1980, 495]. *Proposed by the editors.*

Under what conditions on A can the leading principal $t \times t$ minor of the real orthogonal matrix A be replaced by a symmetric $t \times t$ minor so that the property of orthogonality is preserved?

A solution of this problem appears in this issue (see Solution of E 2840, p. 765).

REVIEWS

EDITED BY J. ARTHUR SEEBACH, JR., AND LYNN A. STEEN
with the assistance of the mathematics departments of St. Olaf, Carleton, and Macalester Colleges
COLLABORATING EDITOR FOR FILMS: SEYMOUR SCHUSTER, CARLETON COLLEGE

The Bicentennial Tribute to American Mathematics: 1776–1976. Edited by Dalton Tarwater. Mathematical Association of America, 1977. vii + 225 pp. $19. (Telegraphic Review, August–September 1978.)

From time to time the MAA has seen fit to encourage the history of mathematics. In this spirit there appeared in 1972 *The Mathematical Association of America: Its First Fifty Years*, edited by the late Kenneth O. May and written by a number of authors, including this reviewer. Five years later the MAA published the volume reviewed here, which grew out of the MAA meeting held in January 1976 to commemorate the bicentennial. As Tarwater writes in his preface, this bicentennial meeting was devoted to "stressing the history of American mathematics. Some speakers were invited to trace American mathematical history from colonial times to the present. Others were selected to address the meeting on various topics of historical interest to the broad mathematical community ... " (p. v). The final eighty pages contain nonhistorical lectures, given at the meeting, which are tangential to the book as a whole and which will not be discussed here.

From its first page this bicentennial volume raises some intriguing methodological questions about the history of mathematics. In his brief essay on mathematics in colonial America (including Canada and Mexico as well), Dirk Struik distinguishes two approaches to the history of mathematics: (1) the "skyline" route which focuses on the great mathematicians and their discoveries, and (2) "the development of mathematics as a social phenomenon as an aid to ... [the] sciences, or as the subject of education" (p. 1). Struik opts for the second approach. Vigorously relating almost three centuries of history in seven pages, he necessarily keeps to the high points. In particular, he recounts how the British influence on colonial mathematics was gradually replaced by French influence after the Revolutionary War.

Many historiographic approaches are possible besides the two which Struik notes. Indeed, the remaining articles encompass a number of these possibilities: biography and reminiscence, as well

as the history of an institution or of a mathematical technology. Yet this list is by no means complete. Even the second article, written by Judith Grabiner, does not fit comfortably into such categories. Among the essayists, she is the sole professional historian of mathematics, a training which becomes evident in her superb treatment of mathematics in the United States from 1776 to about 1900. Largely based on secondary sources, her article synthesizes them into a coherent and readable account which stresses the interplay in nineteenth-century America between mathematics, science, natural theology, Baconian philosophy, governmental institutions, higher education, and economic trends. This is social history of mathematics at its finest.

By contrast, Garrett Birkhoff presents an unusually lengthy article which surveys the leaders in American mathematics from 1891 to the onset of World War II. Although, like Grabiner, he investigates the gradual rise of American mathematicians to international pre-eminence, his article consists mainly of a list of biographical vignettes. Regrettably, it lacks an adequate historical framework in which to place these vignettes, and hence it strikes the reader as fragmented and meandering.

As one draws near to the present, it becomes increasingly difficult to write good history. Nevertheless, the last four essays attempt to do so for the period following World War II. The first of these—a collaborative effort by J. H. Ewing, W. H. Gustafson, P. R. Halmos, S. H. Moolgavkar, W. H. Wheeler, and W. P. Ziemer—employs the "skyline" approach to relate ten significant discoveries of new concepts, examples, methods, or theorems within the preceding three decades. In effect, this lively essay is more exposition than history since the authors devote most of their space to explaining the nature of these discoveries rather than to analyzing their historical development.

On the other hand, the remaining three essays blend historical analysis with a large measure of personal reminiscence. Mina S. Rees recounts her years in the mathematics program of the Office of Naval Research, particularly from 1946 to the founding of the National Science Foundation in 1950. Here she also relies on letters, solicited from a number of those who served with her, to document early governmental support for basic research in mathematics. In what he characterizes as an "impressionistic and personal history," R. W. Hamming describes how computers came to be developed in the United States and how they found uses in number theory, group theory, and differential equations. Lastly, Peter D. Lax discusses the ways in which European émigrés have influenced modern American mathematics.

Beyond its particular content, this bicentennial volume suggests an important question. At a conference on the history of modern mathematics, held in 1974, K. O. May argued for a pluralistic approach to the history of mathematics, whereby both the professional historian and the professional mathematician would contribute their respective expertise (see *Historia Mathematica*, 2 (1975) 449–455). Unfortunately, the present volume reveals how far we remain from such a balanced approach. Except for the article by Grabiner, the research of professional historians of mathematics is nowhere to be seen—even though such historians have become rather numerous. Perhaps this gap reflects the fact that many mathematicians still hold the antiquated view that the history of mathematics can best be done by mathematicians who have become too old to create "real" mathematics. Indeed, in the bicentennial volume S. K. Stein insists that courses in the history of mathematics are "fundamentally guilt offerings" to redeem the inadequate teaching of service courses (p. 197). While this is not the place to do more than remind the reader of the significance of the history of mathematics, the following question is in order: When will the mathematical community recognize that a professional historian of mathematics is as legitimate and useful a component of a mathematics department as is a functional analyst or an algebraic topologist?

GREGORY H. MOORE, University of Toronto

Telegraphic Reviews

Telegraphic reviews are designed to give prompt notice of new books with sufficient information to assist our readers in deciding whether to order an examination copy or to suggest library purchase. Possible uses are indicated as follows:

T = textbook P = professional reading
S = supplementary reading L = undergraduate library purchase
13 to 18 = freshman to second year graduate level usage
1 to 4 = appropriate time in semesters to cover text

Asterisks (*) or question marks (?) denote special positive or negative emphasis, respectively. Publishers are denoted by standard abbreviations; complete addresses may be found in Books in Print.

General, T(15-18: 1), S, P, L*. Measurement for Management Decision. Richard O. Mason, E. Burton Swanson. Addison-Wesley, 1981, xix + 550 pp, $15.95. [ISBN: 0-201-04646-6] A collection of thirty-five nontechnical readings from the literature on the foundations of measurement, and on organizational, individual, and social measurement. Introductory essays and commentary are interspersed for guidance. Each reading is followed by "Questions for Study and Discussion." Valuable and interesting reading for all who receive measurements and base decisions on them. LCL

Precalculus, T(13: 2).** College Algebra and Trigonometry. Bernard Kolman, Arnold Shapiro. Academic Pr, 1981, xvi + 602 pp, $19.95. [ISBN: 0-12-417840-5] A very readable, informal style of presentation including highlighted formulas and definitions, warnings, sample exercises, and chapter summaries, reviews and tests. Includes a chapter on conic sections. Excellent pre-calculus text. JNC

Precalculus, T(13: 1). Fundamental Algebra and Trigonometry, Second Edition. Mervin L. Keedy, Marvin L. Bittinger. Addison-Wesley, 1981, xii + 660 pp, $12.95. [ISBN: 0-201-03839-0] Revisions from First Edition (TR, November 1977) include additions of some exercises and topics. A chapter on probability and statistics has been added, and matrices and determinants are expanded. Nice use of red highlighting. LLK

Education, S(15-16). Activities from the Mathematics Teacher. Ed: Evan M. Maletsky, Christian R. Hirsch. NCTM, 1981, iv + 140 pp, $7 (P). [ISBN:0-87353-173-6] Reprints of thirty-two activities in five categories: Computational Skills, Calculators, Problem Solving, Geometry, and Measurement. Each activity consists of a teachers guide and reproducible student worksheets. Annotated listing of additional "Activities" articles included in each section. Useful for teachers and teacher educators who do not have back issues of The Mathematics Teacher. MW

Education, P, L. Teaching Teachers, Teaching Students: Reflections on Mathematical Education. Ed: Lynn Arthur Steen, Donald J. Albers. Birkhäuser Boston, 1981, x + 136 pp, $12.95. [ISBN: 3-7643-3043-0] This important book highlights issues raised at the August 1980 International Congress of Mathematical Education. First section outlines educational perspectives of five Congress principals; a series of four invited articles follows. A special section includes a potpourri of Congress perspectives and discussion on several Congress themes. The scope is global, the issues raised are current, pertinent, and important. Must reading for teachers of mathematics. PH

Education, P. Guidelines for the Preparation of Teachers of Mathematics: Commission on the Education of Teachers of Mathematics of the National Council of Teachers of Mathematics. NCTM, 1981, vii + 21 pp, $4 (P). [ISBN: 0-87353-177-9] Extensive revision of original Guidelines reflects current thinking on education of mathematics teachers. Lists recommended competencies and suggested courses to develop them for teachers at each level. Note the price for 23 pages of text. MW

Algebra, P, L. Power Series Over Commutative Rings. James W. Brewer. Lect. Notes in Pure and Appl. Math., V. 64. Dekker, 1981, vii + 96 pp, $17.50 (P). [ISBN: 0-8247-6952-X] Presents results, both positive and negative, relating the ideal structure of a ring to that of its power series ring. Accessible to anyone who has had a basic graduate course in algebra. Very readable. JRG

Algebra, T(17-18), S, P. A Lie Group: Rotations in Quantum Mechanics. Jean-Marie Normand. Elsevier North Holland, 1980, xvii + 486 pp, $61. [ISBN: 0-444-86125-4] A pedagogical survey aimed to bring out a set of general methods and results concerning the representations of compact Lie groups, while deriving many specific properties and computational methods for the representations of the three-dimensional rotation group in quantum mechanics. Assumes a general knowledge of quantum mechanics. LCL

Finite Mathematics, T*(13: 1), L. Finite Mathematics with Applications. John J. Costello, Spenser O. Gowdy, Agnes M. Rash. HarBrace J, 1981, xv + 524 pp, $17.95. [ISBN: 0-15-527400-7] Attractively formatted sections introduce mathematical techniques needed in the modeling and solution of problems from business or social science. Techniques include vectors and matrices, linear programming, probability and statistics, computer programming and flow charting. JNC

Calculus, T(14: 1). Introduction to Vector Analysis, Fourth Edition. Harry F. Davis, Arthur David Snider. Allyn & Bacon, 1979, x + 340 pp, $21.55. [ISBN: 0-205-06598-8] Increased applications to physics and engineering, expanded material on line, surface and volume integrals, and general curvilinear coordinates, added supplementary problems, and insertion of "optional reading" sections.

(Third Edition, TR, April 1975.) LLK

Differential Equations, P. Lecture Notes in Mathematics-837: Mathieu Functions and Spheroidal Functions and Their Mathematical Foundation: Further Studies. Josef Meixner, Friedrich W. Schäfke, Gerhard Wolf. Springer-Verlag, 1980, vii + 126 pp, $9.80 (P). [ISBN: 0-387-10282-5] A survey of some of the most important results in the mathematical theory of Mathieu functions and spheroidal functions discovered during the last 25 years. AO

Differential Equations, T(16-17: 1). Perturbation Methods in Applied Mathematics. J. Kevorkian, J.D. Cole. Appl. Math. Sci., V. 34. Springer-Verlag, 1981, x + 558 pp, $42. [ISBN: 0-387-90507-3] Designed as a textbook in applied mathematics. Limit process expansions and multiple-variable expansions are the main tools used to treat singular perturbation problems of the layer and/or cumulative effect type. AO

Differential Equations, T(16-17: 1). Partial Differential Equations. W.E. Williams. Clarendon Pr:Oxford U Pr, 1980, xi + 357 pp, $22.50 (P); $45. [ISBN: 0-19-859633-2; 0-19-859632-4] An introductory textbook emphasizing methods and concepts particularly useful to applied mathematicians (e.g., weak solutions, shocks, Green's functions). AO

Numerical Analysis, T(16-17: 1), P, L. Computer Solution of Large Sparse Positive Definite Systems. Alan George, Joseph W-H Liu. Prentice-Hall, 1981, xii + 324 pp, $24.95. [ISBN: 0-13-165274-5] Presents efficient methods for solving large sparse symmetric positive definite systems of linear equations using a computer. Fortran subroutines implementing many of the algorithms discussed are included. AO

Optimization, P, L.** Queueing Tables and Graphs. Frederick S. Hillier, Oliver S. Yu. Elsevier North Holland, 1981, ix + 231 pp, $32.95. [ISBN: 0-444-00582-X] Extremely thorough presentation of numerical values and graphical comparisons of measures of performance for queueing systems. Both interarrival time and service time distributions may be chosen from the large family of Erlang distributions, including the exponential and degenerate cases. Emphasis on multiple server systems. State of the art. JRG

Geometry, S, P, L*. A Course in Descriptive Geometry. V.O. Gordon, M.A. Sementsov-Ogievskii. Trans: Leonid Levant. MIR Pub, 1980, 376 pp, $10. Invented by the mathematician Monge and long guarded as a French military secret, descriptive geometry has been fundamental to engineering; but its concern with constructing representations of 3-dimensional forms and their use in solving problems now provides insight into the development of computer aided design. Revised from the 1977 Russian edition. JNC

Topology, T(17-18), S, P. Lecture Notes in Mathematics-835: Surfaces and Planar Discontinuous Groups. Trans: J. Stillwell. Springer-Verlag, 1980, ix + 334 pp, $19.50 (P). [ISBN: 0-387-10024-5] Deals with combinatorial topology, combinatorial group theory, and surface theory--much of it developed from scratch. Exercises included. This volume constitutes a revision and expansion of the author's Flächen und ebene diskontinuierliche Gruppen (1970). LCL

Computer Programming, T(13: 1). Business Programming in FORTRAN IV and ANSI FORTRAN 77: A Structured Approach. Asad Khailany. Prentice-Hall, 1981, xvii + 440 pp, $13.95 (P). [ISBN: 0-13-107607-8] Contains basic material on data processing and computer organization along with in-depth coverage of both Fortran IV and ANSI Fortran 77. Uses flow charts. Instructors manual available which includes solutions plus sample exams. LLK

Computer Programming, T(13: 1). Problem Solving and Structured Programming in FORTRAN, Second Edition. Frank L. Friedman, Elliot B. Koffman. Addison-Wesley, 1981, xvi + 514 pp, $11.95. [ISBN: 0-201-02461-6] Emphasizes top down programming. Uses flow diagrams throughout. Incorporates the new features of Fortran 77. Good organization of topics. (First Edition, TR, February 1978.) LLK

Computer Science, T(16-18: 1), S, P, L. Database Security and Integrity. Eduardo B. Fernandez, Rita C. Summers, Christopher Wood. Addison-Wesley, 1981, xiv + 320 pp, $18.95. [ISBN: 0-201-14467-0] Comprehensive coverage and analysis of useful principles and techniques in security (protection of information against unauthorized disclosure, alteration, or destruction) and integrity (correctness of information). Exercises, extensive and up-to-date references and bibliography. Assumes some technical background in computer science. LCL

Reviewers

RJA: Richard J. Allen, St. Olaf; MB: Murray Braden, Macalester; JNC: Judith N. Cederberg, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; JRG: Jennifer R. Galovich, St. Olaf; SG: Steven Galovich, Carleton; JG: Jack Goldfeather, Carleton; PH: Paul Humke, St.Olaf; PJ: Paul Jorgensen, Carleton; LLK: Lorraine L. Keller, St. Olaf; RJK: Roger J. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; JL: Justin Lam, Macalester; LCL: Loren C. Larson, St. Olaf; GHM: George H. Mills, Carleton; AO: Arnold Ostebee, St. Olaf; AWR: A. Wayne Roberts, Macalester; TRS: Thomas R. Savage, St. Olaf; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; SES: Stan E. Seltzer, Carleton; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; FLW: Frank L. Wolf, Carleton.

NEWS AND NOTICES

EDITED BY FRANK KOCHER, The Pennsylvania State University

A CHANGE IN THE MONTHLY

As previously announced, this is the last installment of "News and Notices." In the future personal items and announcements will appear in *Focus*, the newsletter of the Association, as space permits. Newsworthy items should be sent to Marcia Sward, Editor of *Focus*, at the Washington Office of the MAA.

PERSONAL ITEMS

At Northern Arizona University Associate Professor *Lee M. Johnson* has been promoted to Professor of Mathematics; *Michael I. Ratliff* has been appointed Acting Chairman; and *Preben Alsholm*, formerly at the University of Michigan, Dearborn, has been appointed Assistant Professor of Mathematics.

At the Rose-Hulman Institute of Technology, Division of Mathematics, *Lo Yung Su* is on leave at Texas A&M University; *Damon Disch* of Northern Illinois University and *G. Elton Graves* of Idaho State University, have been appointed Assistant Professor; *Brian Winkel*, of Albion College and Michigan Technological University and the founder and editor of *Cryptologia*, has been appointed Associate Professor; and *Gary J. Sherman* has succeeded *William E. Ritter* as Head of the Division.

At the Florida State University Professor *Frederick W. Leysieffer* has been appointed Chairman of the Department of Statistics, where Professor *Morris Skibinski* of the University of Massachusetts and Professor *Robert A. Fontenot* from Whitman College will be spending sabbatical leaves. *Robert Gilmer* of the Department of Mathematics and Computer Science was recently selected as the Robert O. Lawton Distinguished Professor for 1981-82.

At the Metropolitan State College (Denver) *George S. Donovan* and *Ronald D. Whittekin* have been promoted to Professor and *Theophil J. Worosz* has been promoted to Associate Professor.

Rensselaer Polytechnic Institute: *B. David Saunders* of the Department of Mathematical Sciences has been promoted to Associate Professor. Professor *Bobby F. Caviness* has resigned to become Chairman of the Department of Computer Science at the University of Delaware. Professor *Richard C. DiPrima* has resigned as Chairman of the Department of Mathematical Sciences. He has been appointed to the Eliza Ricketts Foundation Chair in Mathematics.

At Western Michigan University *Paul Eenigenberg* and *S. F. Kapoor* have been promoted to the rank of Professor.

New England College: Professor *Eric C. Nummela* has been named Coordinator of the Division of Science and Engineering. Professor *Edward T. Ordman*, on leave of absence for this academic year, is Associate Professor of Mathematics at Memphis State University.

Professor *Benjamin Epstein* of the Technion-Israel Institute of Technology, received the 1981 Reliability Society Award of the Institute of Electrical and Electronic Engineers for "laying the basic foundation of the statistical theory of life testing and for the development of its mathematical models.

Associate Professor *Stephen H. Friedberg* of the Illinois State University has been promoted to Professor. He is spending the current academic year at the University of Missouri.

Professor *Walter J. Hendricks* of Case Western Reserve University is spending the 1981-82 academic year as a visitor in the Mathematics Department of the University of Virginia.

Professor *Edwin T. Hoefer* of Rochester Institute of Technology will be teaching at the Polytechnic of Wales for 1981-82 academic year. He will be part of the Fulbright Exchange Program administered through the U.S. Office of Education.

Associate Professor *James T. Lewis* of the University of Rhode Island has been promoted to Professor.

Associate Professor *Frank W. Owens* of Ball State University has been promoted to Professor.

Professor *Dan Pedoe* of the University of Minnesota has retired recently with the rank of Professor Emeritus. He has been a member of the School of Mathematics since 1965.

David E. Zitarelli of Temple University has been appointed Director of the Outstanding Achievement Scholarship program at that University.

Herbert R. Leifer of Pittsburgh, PA, died May 14, 1981. He has been a member of the Association since 1934.

John W. Jewett, Regents Professor at Oklahoma State University, died July 5, 1981 at the age of 52. He was a member of the Association for twenty years.

The deaths of these members of the Association have been recently reported: *Martin Helling* of Youngstown State University; *William H. Landis* of Martinez, CA; and *Milton A. Nelson* of Great Neck, N.Y.

ERRATA

In the October 1981 issue of the MONTHLY, page 632, *Lawrence A. Salcman* should be listed as *Lawrence A. Zalcman*.

On page 634 Professor *Jean J. Pederson* should be Professor Jean J. Pedersen.

On page 635, The Association for Women in Mathematics is incorrectly called The Association of Women in Mathematics.

MATHEMATICAL ASSOCIATION OF AMERICA

Official Reports and Communications

SPRING MEETING OF THE TEXAS SECTION

The Annual Spring Meeting of the Texas Section was held at San Antonio College in San Antonio, Texas on April 10-11, 1981. Professor *Ray Tebbets* of San Antonio College was in charge of arrangements. Registered attendance was 185.

Invited speakers included Professor *Larry Schumaker* of Texas A&M University who spoke on "Fitting Surfaces to Scattered Data;" Professor *Charles David Miller* of American River College in Sacramento, California, who reviewed the current status of Mathematics Laboratories; Dr. *Preston D. Kronkosky*, Deputy Executive Director of the Southwest Educational Development Laboratory, who presented a review and comments on the White House Report on "Science and Engineering Education for the 1980's and Beyond;" and Professor *John Jobe* of Oklahoma State University (representing the MAA) who presented a Career Awareness Package for Secondary School Students titled "Mathematics at Work in Society."

A discussion of the status of mathematical research and graduate programs in Texas was presented by Professors *Jim Daniel*, Chairman of Mathematics at the University of Texas at Austin; *John Ed Allen* Chairman of Mathematics at North Texas State University; *Elton Lacey*, Chairman of Mathematics at Texas A&M University; *Bob Thrall* of Rice University; and *Dalton Tarwater* of Texas A&M University.

A panel discussion of "Remedial/Developmental Mathematics in Colleges and Universities" was presented by Professors *Vic Morgan* of Sul Ross State University, *Richard Alo* of Lamar University, *Douglas Tharp* of the University of Houston--Downtown College, and *Bennie Zinn* of San Antonio College. Professor Morgan reviewed Remedial/Developmental Mathematics in Colleges across the country and other panelists detailed practices at representative institutions within the state.

Professor *John Ed Allen* of North Texas State University gave a brief report on the work of the Ad Hoc Committee of the Texas Association of Academic Administrators in the Mathematical Sciences which is continuing its work on monitoring and assisting in the process of revising certification requirements for school mathematics teachers in Texas. Significant was information that this group has had some opportunity for input in determining standards.

A progress report on the revitalization of the Texas Section High School Lecture Program was presented by *William V. McNabb* of Skyline Center, Director of this program.

A special session for student papers was incorporated into this year's program. Five papers were presented. Professor *Larry Heath* of the University of Texas at Arlington made the arrangements.

Student honorees at the Section Banquet this year were Texas Mathematical Olympiad winner *Gerhard Paseman* of Bellaire High School in Houston and outstanding performer *Jerome de La Cruz* of John Marshall High School in Santonio. Also honored in a special tribute was Professor *Robert Greenwood* of The University of Texas at Austin for distinguished service to mathematics and to the Texas Section of MAA. Professor Greenwood entertained with a first hand account of some of the early history of the Texas Section and closed his response with one of his famous "magic" acts. He is retiring at the end of the current school year.

The Spring Meeting incorporated special group meetings for department chairmen. (Texas Association of Academic Administrators in the Mathematical Sciences), the Mathematical Association of Two-Year Colleges (TexMATYC), Institutional Representatives of the Texas Section and the Executive Committee. Officers and directors for 1981-82 are as follows: Chair, *Robert L. Tennison*, University of Texas at Arlington; Sec./Treas., *Glen Mattingly*, Sam Houston State University; Chair-Elect, *John Ed Allen*, North Texas State University; Immediate Past Chair, *Bill D. Anderson*, East Texas State University; Level-I Director, *Vivian A. Dennis*, Eastfield College; Level-II Director, *R. Vic Morgan*, Sul Ross State University; Level-III Director, *Bennie B. Williams*, University of Texas at Arlington; Director-at-Large, *Landon Colquitt*, Texas Christian University; Arrangements Chair, *Richard A. Alo*, Lamar University; Arrangements Chair-Elect, *Melvin R. Hagan*, North Texas State University; Texas AHSM Director, *J.R. Boone*, Texas A&M University.

Contributed papers were as follows: Faculty papers: "The Left Singular Ideal in Semiprime Rings," *Baxter Johns*, Baylor University; "Some Properties of Valuation Rings (Expository)," *Nick Vaughan*, North Texas State University; "A Geometric View of the Equation $a^2 + ab + b^2 = 3c^2$," *George Berzsenyi*, Lamar University; "Why Are the Times 8:18 and 10:09 so Pleasant?" *Montie G. Monzingo*, Southern Methodist University; "More About the Mathematical Aspects of a Lunar Shuttle Landing," *Bill Anderson* and *John Lamb*, Jr., East Texas State University; "Archimedes' Mind," *Ali R. Amir-Moez*, Texas Tech University; "Analysis of the Effect of Electromagnetic Pulses on Penetrable Bodies," *David Cohoon*, School of Aerospace Medicine, Brooks A.F. Base; "Distribution: A 'Real' Approach," *James A. Bell*, Laredo State University; "Statistical Technique Selection Models," *Chris Boldt*, Eastfield College; "College Mathematics: Four Points of View," *Ralph W. Cain*, The University of Texas at Austin; "Applied Mathematical Sciences: An Undergraduate Program for the 80's," *Michael G. Murphy*, University of Houston--Downtown; "A Mathematics Correspondence Course," *M.K. Jones*; "The Interaction of Student Characteristics & Instructional Style," *Patricia S. Hickey*, Baylor University; "Cartoons in the Classroom," *Joe F. Allison*, Eastfield College; "Semi-Regular and Semi-Normal Spaces," *Charles Dorsett*, Texas A&M University; "The Z Lattice of Upper Semi Continuous Functions and Compact Hausdorff Spaces," *Don E. Edmondson*, The University of Texas at Austin; "P.L. Approximate Fibrations," *Robert E. Goad*, Sam Houston State University; "Constable Cellular Decompositions of S^3," *Richard T. Denman*; "Quasi Expanding Maps and Fixed Points," *B.B. Williams* (with *A.A. Gillespie*), The University of Austin at Arlington; "Differentiability of Infinite Dimensional Curves," *Russell Bilyeu*, North Texas State University; "The Monodiffric Laplace Transform," *Tahereh Daneshi/C.R. Deeter*; "Steady State Solutions of Differential Equations and the Laplace Transform," *Jean Richmond*, Southern Methodist University; "A Characteristic of the Characteristics Function," *Fred Curtis*, The University of Dallas; "Deriving Distribution Functions for Characteristic Functions," *Larry Ketchersid*, Trinity University; "Generalized Circular and Hyperbolic Functions," *Craig Finley*, East Texas State University; "Physics and Mathematical Modeling," *Jim Armstrong*, The University of Texas at Arlington; "The Game of Blockade," *Jeffrey A. Conly*, The University of Texas at Arlington.

THE MATHEMATICAL ASSOCIATION OF AMERICA
THE SIXTY-FIRST SUMMER MEETING OF THE ASSOCIATION

The Sixty-First Summer Meeting was held at the University of Pittsburgh in the period August 17-19, 1981. There were 785 registrants including 534 members of the Association. The meeting was held in conjunction with meetings of American Mathematical Society, Association for Women in Mathematics and Pi Mu Epsilon.

Sessions of the Association were held in David Lawrence Hall and the Law School. The Program Committee consisted of W. E. Deskins, Chairman; William A. Beck, Allan G. Bluman, John D. Bradburn, Charles A. Cable, Barbara T. Faires, Frank Hergeist, Beverly Michael, Earle F. Myers, John O. Riedl, Jr., and Melvin Woodard. The Program was comprised of the following presentations:

FIRST SESSION OF THE ASSOCIATION

The Earle Raymond Hedrick Lectures: "Finite Simple Groups," Daniel Gorenstein, Rutgers University. Lecture I: "The Enormous Theorem"

The First Lecture gave a broad overview of the classification theorem, including an explanation of the extreme length of its proof.

"The Trouble with Area," Marvin Knopp, Temple University.

This was an expository talk dealing with the genuine annoying difficulties that arise in any attempt to develop rigorously the theory of area (or Jordan content). Exactly where the difficulties lie depends upon the specific choice among the many possible definitions of inner and outer area, but they persist in any case, a circumstance which explains why a rigorous treatment is not given in the standard calculus sequence. Topics discussed included: invariance of area under rigid motions; the connection with the Riemann integral; the connection with plane topology; area of unbounded sets and the Cauchy-Riemann integral.

"Development of the Theory of Cluster Sets," A.J. Lohwater, Case Western Reserve.

The development of the theory of cluster sets was traced from the Casorati-Weierstrass theorem on the limiting values of an analytic function in the vicinity of an isolated essential singularity. The notation of Seidel has influenced the applicabilty of the theory and some of the modern applications were cited.

SECOND SESSION OF THE ASSOCIATION

Hedrick Lecture II: "The Friendly Giant and His Relatives," Professor Gorenstein.

The second lecture discussed the sporadic groups, which are not part of any infinite family of simple groups, the largest of these being Griess' "friendly giant". This lecture described the group-theoretic origins of the sporadic groups.

"Constant-weight Codes, Sum-free Sets and Harmonious Graphs," Ronald L. Graham, Bell Telephone Labs.

A fundamental problem in coding theory is the estimation of $A(n,d)$, the maximum number of binary strings of length n one can have so that any two strings differ in at least d places. Very recently, new techniques have been discovered which have greatly improved previous estimates of $A(n,d)$, especially in the case when all the binary strings are required to have the same number of ones. This talk described these developments and showed how they are related to classical questions in combinatorial number theory as well as new questions in graph theory.

"Operations Research in the Federal Reserve System," Patrick L. Hayes, Federal Reserve System.

Presented were three mathematical models recently developed to aid in analyses of operational problems in the Federal Reserve System. Designed to introduce undergraduates, as well as faculty, to the operations research-modeling approach, this talk discussed optimization, time series, and casual models currently applied to determine more efficient resource allocations. Emphasis was placed on aspects of "real world" modeling.

"Combinatorics and Geometry," D. K. Ray-Chaudhuri, Ohio State University.

Panel Discussion: Institutional Responses to "Math Anxiety," David A. Blaeuer, SUNY, College at Buffalo, Clifford A. Baylis, Jr., Community College of Allegheny County, Rosalie B. Jackson, Waynesboro College, and Beverly K. Michael, University of Pittsburgh.

To reduce tension and anxiety among students in mathematics classrooms, structure the course so as to include personal contact in groups of up to four. This contact should take the form of activities of several kinds, including focusing on individual feelings and attitudes about many areas including mathematics, and on solving mathematics problems. This should happen in a supportive environment where much learning results from student-student interactions, from a commitment to a group and from instructor explanations and comments. The goal is more fully functioning human beings who just happen to be increasing their mathematical knowledge.

Open Session: Programs in Mathematics in Four-Year Colleges and Universities, Barnet M. Weinstock, UNCC and Gail S. Young, Case Western Reserve University.

THIRD SESSION OF THE ASSOCIATION

Hedrick Lecture III: "The Thirty-Years War," Professor Gorenstein.

The classification theorem was proved over a roughly thirty year period and involved the combined efforts of more than a hundred mathematicians. In the third lecture the highlights of some of the many battles this effort entailed were given.

"The Math Workshop", Deborah Hughes-Hallett, Harvard University.

"Transition from Academia to Industry," Erin Cramer, Aerospace Corporation.

MAA Business Meeting: Presentation of the Allendoerfer, Ford and Polya Awards for MAA-published exposition.

FOURTH SESSION OF THE ASSOCIATION

"The Interaction of Complex-Analytic Geometry and Theoretical Physics," R. O. Wells, Jr., Rice Univ.

In the past 15 years a relationship between the geometry of complex-analytic manifolds and the solutions to some of the fundamental differential equations describing phenomena in elementary particle theory and in relativity theory has emerged. This theory, called "twistor theory," after its initiator, Roger Penrose, has had a remarkable success in relating problems in mathematical physics to sometimes tractable problems in algebraic geometry and several complex variables. This lecture described these recent developments.

"Statistics and the Law," Mary W. Gray, American University.

The law's reliance on probalistic models, while pervasive, is usually not recognized, much less understood, by lawyers, judges or juries. Instances of the use and misuse of statistical evidence were presented, ranging from Poincare's defense of Dreyfus to recent United States Supreme Court findings of discrimination. Alternative approaches to "proof by numbers" were investigated.

"On the Way Primes, Sums of Squares and Other Integer Sequences are Distributed," Heini Halberstam, University of Illinois.

The object of the talk was to show what can be proved by elementary methods about the distribution of primes, of numbers that are sums of two squares, of squarefree integers, and of other integer sequences. The talk described also the current state of knowledge about these topics.

"Community College--Fast-Paced Mathematics," John R. Starmack, Community College of Allegheny County.

South Campus, in cooperation with John Hopkins University (Office of Talent Identification and Development) and the Allegheny Intermediate Unit, offers a Fast-Paced Mathematics Program. Pre-calculus mathematics courses are taught to a homogeneous group of gifted students who reason exceptionally well mathematically. Successful completion was defined as scoring at or above the 85th percentile on the Cooperative Mathematics Test Series designed by E.T.S. This program should be viewed as an opportunity to learn at a rate commensurate with a student's ability and an option available to better serve talented students.

"About That Text You are Planning to Write," Marvin Schlichting, Triton College.

The speaker commented on his own experiences in writing texts for Basic Algebra and Intermediate Algebra. He discussed contacts with publishers, critical reviews, planning, proofreading, graphics, time and rewards.

"Astrophysics and Cosmology," Cyril Hazard, University of Pittsburgh.

SPECIAL SESSIONS OF THE ASSOCIATION

The Association's Film Program was presented on Monday evening and included showings of the following: Graphs of Complex Functions: w=z and w=e; Shapes of the Future: Some Unsolved Problems in Geometry--Two Dimensions; The Seven Bridges of Konigsberg; Adventures in Perception; Modeling Surveys (BBC broadcast--part of the Open University's Foundation Course in Mathematics); Powers of Ten; Symmetries of the Cube; Circle Circus; Math Anxiety: We Beat It, So Can You!

Professor Ronald Wenger of the University of Delaware organized a special session entitled "Computers in the Undergraduate Mathematics Curriculum" held on Tuesday evening. The following is a list of participants:

"Introduction to the Session and Report on the National Consortium", Ronald H. Wenger, U. Delaware; "Instructional Software in the Undergraduate Math Curriculum", Theron D. Rockhill, SUC, Brockport; "Use of CONDUIT Materials in Teaching Mathematics", David A. Smith, Duke University; "Results of a Project to Incorporate the Computer into the Mathematics Curriculum," Sheldon P. Gordon, Suffolk Community College; "Simulation and Mathematical Modeling Using Computers," Marialuisa McAllister, Moravian College.

Professor Gerald J. Porter of the University of Pennsylvania organized the Wednesday evening session "Microcomputer Graphics in Undergraduate Mathematics." The other participants were Mark John Christensen, Georgia Tech; Roy E. Myers, Penn State at New Kensington. An abstract follows:

Previous sessions have shown the uses of computer graphics in undergraduate mathematics instruction. Most of the earlier work involved the use of large computers and expensive graphics terminals. It was the purpose of this session to demonstrate the variety of applications which can be done on inexpensive microcomputers.

The Association presented another of its sequence of Mini-Courses. This course was arranged by the ad hoc Committee on Continuing Education chaired by John O. Riedl, Jr. of Ohio State University. There were 46 registrants for this mini-course the abstract of which follows:

MAA MINI-COURSE: "The Use of Computers to Teach Mathematics," Donald O. Norris, Ohio University.

A brief introduction to BASIC programming was given followed by discussions of how computers can be used in a variety of courses. This included the more traditional approaches such as the use of computers in calculus and differential equations courses as well as the use of simulation models in liberal arts courses or mathematics education classes. Microcomputers were available for the use of the participants.

The Association sponsored a Tuesday evening reception and dinner in honor of twenty-five year members. There were approximately 70 in attendance. The member present for the evening with the longest record of MAA membership was Wyman L. Williams who has been a member since 1924.

ENTERTAINMENT

On Wednesday evening, there was a reception sponsored by the Allegheny Mountain Section of MAA. The traditional Summer Meeting Picnic was held on Thursday evening at the Athletic Shelter in Schenley Park. There were tours of Fallingwater, the famous summer home of the Kauffman family designed by Frank Lloyd Wright, on both Tuesday and Thursday. Many members of the Association attended one or more of the baseball games held between the Pittsburgh Pirates and the San Francisco Giants during the time of the meetings.

Robert Sutherland Lord, organist and faculty member at the University of Pittsburgh, gave a concert at Heinz Memorial Chapel on Wednesday at 12:15 P. M.

The membership of the Local Arrangements Committee was Jacob Burbea, Chairman; Earle F. Myers, Publicity Director; Elayne Arrington-Idowu, F. Gonzalez Asenjo, William A. Beck, Mario Benedicty, Frank T. Birtel (ex-officio), Henry W. Block, W. Eugene Deskins, Barbara T. Faires, James P. Fink, William G. Fleissner, Ka-Sing Lau, William J. LeVeque (ex-officio), David P. Roselle (ex-officio), Kathleen Ann Taylor, Earl G. Whitehead, and Melvin Woodard.

MEETING OF THE BOARD OF GOVERNORS

The Board of Governors met at 9:00 A.M. on Sunday, August 16, in Forbes Quadrangle with 38 members in attendance. Among the items of business approved and items discussed at this meeting were the following:

The Board elected a slate of Associate Editors of the MONTHLY proposed by Professor Paul R. Halmos, who begins his term as Editor on January 1, 1982. The term of office of the Associate Editors, listed below, is through the end of 1986. Those elected are Thomas F. Banchoff, Brown University; Ralph P. Boas, Northwestern University; David Borwein, University of Western Ontario; Joel L. Brenner, Palo Alto, CA; Vladimir Drobot, University of Santa Clara; John H. Ewing, Indiana University; Richard K. Guy, University of Calgary; Deborah T. Haimo, University of Missouri, St. Louis; Franklin T. Haimo, Washington University; Seymour Schuster, Carleton College; J. Arthur Seebach, Jr., St. Olaf College; Lynn A. Steen, St. Olaf College; Mary R. Wardrop, Central Michigan University; Robert F. Wardrop, Central Michigan University; Herbert S. Wilf, University of Pennsylvania. Also elected for the one-year term extending through December 31, 1981, was Allan L. Edmonds of Indiana University.

The Board accepted the following grants:

1. \$6000 from Connecticut Mutual Life Insurance Company for support of the Hartford Region of the secondary school lectureship program entitled "Blacks and Mathematics."

2. \$9500 from the International Business Machines Corporation for support of the Awards Ceremony in honor of the winners of the 1981 USA Mathematical Olympiad.

3. \$10,900 form the National Science Foundation for continuation of the MAA-sponsored project entitled "Mathematics at Work in Society."

4. \$1000 from the Polaroid Foundation for support of the Boston Region of the secondary school lectureship program entitled "Women and Mathematics."

The Board voted to appoint a standing Committee on Mini-Courses. It is intended to continue offering such courses at national meetings and to begin offering them in conjunction with Section meetings, as well.

The Board heard the report from the Committee on the Undergraduate Program in Mathematics that its new description of a general mathematical sciences curriculum will be available during Fall, 1981. Copy of the publication will be mailed free of charge to all department chairmen and additional copies will be available for sale from the Association's headquarters. Preparation and publication were supported by a grant from the Sloan Foundation.

The Board approved appointment of a Joint (AMS-MAA) Program Committee for the August, 1982, meeting. This marks the first attempt to integrate the scientific programs of the two organizations.

A new membership option in which MATHEMATICS MAGAZINE can be the journal selected was approved. This new option will first be available in 1983.

The Board approved meeting at the State University of New York at Albany in August, 1983. With this, the schedule of meetings now scheduled is as follows:

Cincinnati	January 13-17, 1982
Toronto	August 23-27, 1982
Denver	January 5-9, 1983
SUNY Albany	August 8-12, 1983
Louisville	January 25-29, 1984
Anaheim	January 9-13, 1985
San Antonio	January 21-25, 1987

A compendium of the media articles about the International Mathematical Olympiad was distributed to the Board members and Professor Henry L. Alder presented his report as Chairman of the Commission on the International Mathematical Olympiad. Interested persons can obtain either from Professor Alder (University of California, Davis, CA 95616).

It was reported by MAA Associate Director Marcia Sward that the twenty-year members of the Association have made contributions to "Friends of FOCUS" totalling more than $4200. Dr. Sward also described some of the plans she has as Editor for future issues of the Association's attractive new newsletter, FOCUS.

Professor Leonard Gillman presented his report as Treasurer. He reported that the recently acquired headquarters continues to be a sensational investment in that it allows the Association to have minimal operating expenses while providing rental income and capital appreciation. The Treasurer also reported that the Association had essentially broken even during 1980. Finally, the Board approved an operating policy for the Committee on Investments.

President Anderson led a discussion of the budgetary cuts that seem to be in store for the Education Directorate of NSF. Following this discussion, the Board adopted the following resolution: "The Board of Governors of the Mathematical Association of America expresses its grave concern over the proposed cuts in funding for science education. President Anderson is instructed to convey this concern to appropriate members of Congress and the Reagan administration."

The Board of Governors approved the following amendment to the By-Laws proposed for adoption by the membership at the Business Meeting of the Association at the January, 1982 meeting of the Association:

PROPOSED AMENDMENT (ARTICLE IV PARAGRAPH 8)

> The Association shall, to the extent allowed by law, indemnify any member or former member against expenses (including attorney's fees), judgements, fines and amounts paid in settlement actually and reasonably incurred by such member in connection with the defense of any action, suit, or proceeding in which he or she is made a party by reason of having been a member and having acted in good faith as a member of the Board of Governors or other committee. The Association may purchase and maintain insurance on behalf of such member to provide for indemnifying him or her against such liabilities.

MEETING OF SECTION OFFICERS

On Sunday evening at 7:00 P.M. there was a meeting of Section Officers. The theme of the meeting was "How to Make Section Meetings More Attractive." Persons who wish copy of the minutes of this meeting or who have suggestions to contribute are invited to correspond with the Chairman of the Committee on Sections, S. W. Hahn (Wittenberg University, Springfield, OH 45501).

ELECTIONS TO MEMBERSHIP

The Board of Governors elected the following as institutional members of the Association:

Ball State University
Farleigh Dickinson University
The Lindenwood Colleges
Rollins College
University of Baltimore
University of Pennsylvania
University of Wisconsin-LaCrosse
Centro de Investigacion en Matematics
Glassboro State College
Schoolcraft College
Southern Connecticut State College
University of Miami
University of Southern California
University of Wisconsin-Stevens Point

David P. Roselle, Secretary

CALENDAR OF FUTURE MEETINGS

Sixty-fifth Annual Meeting, Cincinnati, Ohio, January 15–17, 1982.
Sixty-second Summer Meeting, Toronto, Canada, August 23-25, 1982.

The following is a list of the Sections of the Association with dates of future meetings so far as they have been reported to the Editorial Director.

ALLEGHENY MOUNTAIN, Allegheny College, Meadville, Pennsylvania, April 1982.
EASTERN PENNSYLVANIA AND DELAWARE, Saturday before Thanksgiving.
FLORIDA, Valencia Community College, Orlando, March 5-6, 1982.
ILLINOIS, Southern Illinois University, Edwardsville, April 30–May 1, 1982.
INDIANA
INTERMOUNTAIN
IOWA, Grinnell College, Grinnell, March 26-27, 1982.
KANSAS, Emporia State University, Emporia, April 2–3, 1982.
KENTUCKY, University of Kentucky, Lexington, April 2-3, 1982.
LOUISIANA-MISSISSIPPI, University of Southwestern Louisiana, Lafayette, February 12-13, 1982.
MARYLAND–DISTRICT OF COLUMBIA–VIRGINIA, Saturday before Thanksgiving and last Saturday in April.
METROPOLITAN NEW YORK, spring. Deadline for papers two weeks before meeting.
MICHIGAN, first Friday and Saturday in May. Deadline for papers six weeks before meeting.
MISSOURI, University of Missouri, Rolla, April 9–10, 1982.
NEBRASKA, Kearney State College, Kearney, April 2–3, 1982.
NEW JERSEY, early November and early May.
NORTH CENTRAL, end of October and April. Deadline for papers October 1 and April 1.
NORTHEASTERN, Saturday before Thanksgiving and third week in June.
NORTHERN CALIFORNIA, University of California, Davis, February 20, 1982.
OHIO, Capital University, Columbus, April 30–May 1, 1982.
OKLAHOMA–ARKANSAS, University of Arkansas, Fayetteville, March 25–27, 1982.
PACIFIC NORTHWEST, Western Washington University, Bellingham, Washington, June 17–19, 1982.
ROCKY MOUNTAIN, Western State College, Gunnison, Colorado, April 30–May 1, 1982.
SEAWAY, first Saturday in November and Saturday in late April. Deadline for papers six weeks before meeting.
SOUTHEASTERN, Emory University, Atlanta, Georgia, April 1982.
SOUTHERN CALIFORNIA, first or second Saturday in March.
SOUTHWESTERN, University of Arizona, Tucson, April 1982.
TEXAS, Lamar University, Beaumont, April 9–10, 1982.
WISCONSIN, University of Wisconsin, Fond du Lac, late March 1982.

FUTURE MEETINGS OF OTHER ORGANIZATIONS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Washington, D.C., January 3–8, 1982.
AMERICAN MATHEMATICAL ASSOCIATION OF TWO YEAR COLLEGES, Las Vegas, Nevada, November 11–14, 1982.
AMERICAN MATHEMATICAL SOCIETY, Cincinnati, Ohio, January 13–16, 1982.
AMERICAN SOCIETY FOR ENGINEERING EDUCATION
ASSOCIATION FOR COMPUTING MACHINERY
ASSOCIATION FOR SYMBOLIC LOGIC, Philadelphia, Pennsylvania, December 1981.
ASSOCIATION FOR WOMEN IN MATHEMATICS, Cincinnati, Ohio, January 13–17, 1982.
CANADIAN SOCIETY FOR HISTORY AND PHILOSOPHY OF MATHEMATICS/SOCIÉTÉ CANADIENNE D'HISTOIRE ET DE PHILOSOPHIE DES MATHÉMATIQUES
FIBONACCI ASSOCIATION
INSTITUTE OF MATHEMATICAL STATISTICS
MU ALPHA THETA
NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS, Toronto, Ontario, Canada, April 14–17, 1982.
OPERATIONS RESEARCH SOCIETY OF AMERICA, Detroit Plaza, Detroit, Michigan, April 19-21, 1982.
PI MU EPSILON
SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION
SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS, Stanford University, Stanford, California, July 19–23, 1982 (30th Anniversary Meeting).

ERRATA

Vol. 87

Solomon W. Golomb, "Iterated Binomial Coefficients" (pp. 719–727):

Page 721: In line 7, read $(n-k)(n+k-1)$ instead of $(n+k)(n-k-1)$.
In line 9, in the displayed formula, read

$$\binom{k}{2} \text{ instead of } \binom{n}{k} \text{ (as in the display in line 1).}$$

In line 11, read $\Pi_{j=-k}^{k-1}(n+j)$.

Page 726: In line 11, in the binomial coefficient, read 2 for a.

Vol. 88

James G. Kennedy, "Arithmetic with Roman Numerals" (pp. 29–32): The author wishes to make it clear that he did not claim to have invented the multiplication algorithm; in fact, in his original manuscript he remarked that he had seen it in an article in *Nature* but could not find the reference. That reference has now been supplied by H. L. Armstrong: Margaret Lazarides, Quare multiplicandum est, *Nature*, 226 (1970) 195.

Saunders Mac Lane, "Award for Distinguished Service to Professor Ralph P. Boas, Jr." (pp. 85–86):

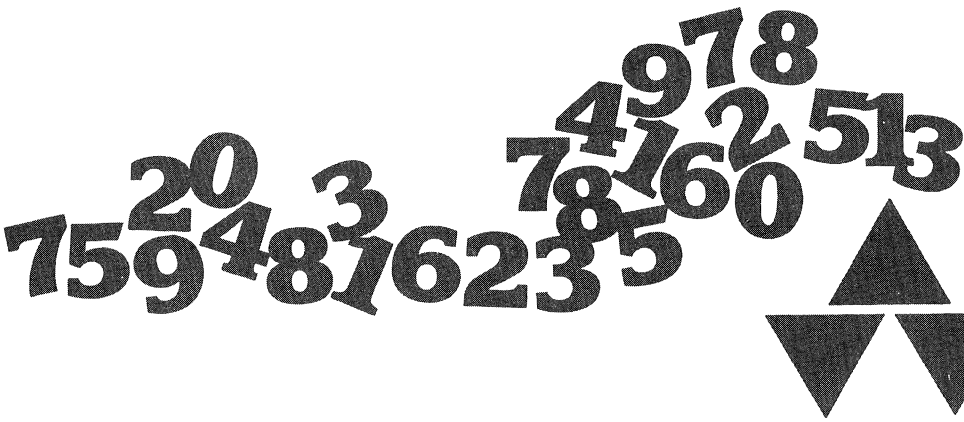
Page 85: In line 21, for "first" read "second." (The first mathematician employed full-time by the A.M.S. was Dr. Caroline Eustice Seely, from 1914 to 1936.)

Ezra Brown, "The First Proof of the Quadratic Reciprocity Law, Revisited" (pp. 257–264): The following line was omitted from the top of page 259:

"LEMMA 2. *If* $(a, p) = 1$, *then* $a^{(p-1)/2} \equiv -(a \mid p)(p-1)! \pmod{p}$."

Donald E. Knuth, "A Permanent Inequality" (pp. 731–740). Page 731:

Note added in proof: It was recently learned that part of Egorychev's result was anticipated a year earlier by D. I. Falikman, whose elegant proof appears in *Matematicheskie Zametki* 29 (June 1981), 931–938. Falikman's paper, which was received for publication on May 14, 1979, establishes the minimal value of doubly stochastic permanents but does not show that this value is uniquely attained.



and in Calculus Texts.

Calculus, Third Edition
Lynn H. Loomis, Harvard University
0-201-05045-5, Hard, 1,000 pp. (tent.), available February 1982
Student Supplement and Solutions Manual available.

Calculus With Applications
Daniel L. Auvil, Kent State University, Stark Campus
0-201-10063-0, Hard, 448 pp. (tent.),
available February 1982
Student Supplement and Solutions Manual
available.

Mathematics: A Modeling Approach
Marvin L. Bittinger, Indiana University-
Purdue University at Indianapolis
J. Conrad Crown, Indiana University-
Purdue University at Indianapolis
0-201-03116-7, Hard, 720 pp., available now
Instructor's Manual available.

To receive a complimentary copy of any
of these books, please write to:

Marianne C. P. Rutter
College Marketing
Department MM
Addison-Wesley Publishing Company, Inc.
Reading, Massachusetts 01867

Science & Mathematics Division
Addison-Wesley Publishing Company, Inc.
Reading, Massachusetts 01867

CONTENTS

Manuscripts of main articles should now be submitted to P. R. Halmos, Department of Mathematics, Indiana University, Bloomington, IN 47405.

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.